

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:01:28 ; Search time 5.29303 Seconds
(without alignments)
817.488 Million cell updates/sec

Title: US-09-986-682B-1
Perfect score: 114
Sequence: 1 MNSGDYKEDYGFATHTRADML 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPTRMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_protent:*
13: sp_virus:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	100.0	487	2	082854
2	59	51.8	489	16	097181
3	55	48.2	487	10	093290
4	55	45.6	646	10	092TV1
5	50	43.9	524	3	09P330
6	50	42.1	478	16	09Y0A8
7	48	42.1	499	2	09KRS8
8	48	42.1	582	4	0925E5
9	48	42.1	670	4	08WV3
10	48	42.1	673	4	09BRR2
11	48	42.1	832	11	09CXR2
12	47	41.2	289	10	09S2J8
13	47	41.2	435	10	065620
14	47	41.2	567	16	098R03
15	46	40.4	365	17	059328
16					093288

17	46	40.4	677	10	092051	092951 arabisdopsis
18	45	39.5	201	5	09VTV3	09vtv3 drosophila
19	45	39.5	526	3	096TV7	096tv7 venturia na
20	45	39.5	608	16	09H2V4	09h2v4 pseudomonas
21	45	39.5	649	13	09DDC0	09ddc0 pleurodeles
22	45	39.5	728	10	004512	004512 arabisdopsis
23	45	39.5	729	10	092U22	092u22 arabisdopsis
24	45	39.5	749	10	09C6M5	09c6m5 arabisdopsis
25	45	39.5	882	10	065231	065231 arabisdopsis
26	45	39.5	893	10	09LHT1	09lht1 arabisdopsis
27	45	39.5	893	10	080973	080973 arabisdopsis
28	45	39.5	941	10	09LTH0	09lth0 arabisdopsis
29	45	39.5	942	10	09SLI8	09sl18 arabisdopsis
30	45	39.5	942	10	092U08	092u08 arabisdopsis
31	45	39.5	943	10	09SFY7	09sfy7 arabisdopsis
32	45	39.5	1071	10	09MAK2	09mak2 arabisdopsis
33	45	39.5	1127	10	09C608	09c608 arabisdopsis
34	45	39.5	6781	12	091AV2	091av2 porcine epi
35	44	38.6	210	2	08RIW9	08riw9 campylobact
36	44	38.6	250	16	09A735	09a735 caulobacter
37	44	38.6	253	12	09PXJ7	09pxj7 transmissib
38	44	38.6	342	4	09HCA5	09hca5 homo sapien
39	44	38.6	364	5	021916	021916 caenorhabdi
40	44	38.6	365	3	09HDU4	09hdu4 schizosacch
41	44	38.6	392	3	096WU4	096wu4 zygosacchar
42	44	38.6	459	17	08U3P5	08u3p5 pyrococcus
43	44	38.6	484	16	09PGI8	09pgi8 xylella fas
44	44	38.6	600	3	09HG12	09hg12 neurospora
45	44	38.6	677	13	090XA5	090xa5 cyprinus ca

ALIGNMENTS

RESULT 1

ID	082854	PRELIMINARY:	PRT:	487 AA.
AC	082854;			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	Beta-fructofuranosidase.			
OS	Bacillus sp.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;			
OC	Bacillaceae; Bacillus.			
OX	NCBI_TaxID-1409;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-V230;			
RA	Tsutsaki K.;			
RT	"Cloning and sequencing of beta-fructofuranosidase gene from Bacillus			
RT	sp. V230";			
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-V230;			
RA	Kurimoto M.;			
RT	"Cloning and sequencing of beta-fructofuranosidase gene from Bacillus			
RT	sp. V230";			
DR	EMBL; AB010272; BAA32083.1; -			
DR	InterPro; IPR003469; Glyco_hydro_68.			
DR	Pfam; PF02435; Glyco_hydro_68; 1.			
SQ	SEQUENCE 487 AA; 53412 MW; 7F52A47921824AE3 CRC64;			

Query Match 100.0%; Score 114; DB 2; Length 487;
Best Local Similarity 100.0%; Pred. NO. 2.2e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFATHTRADML 21
|||||
DB 33 MNSGDYKEDYGFATHTRADML 53

RESULT 2

097181
ID 097181 PRELIMINARY: PRT: 489 AA.
AC 097181:
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN Levanusucrase.
OS CAC1772.
OC Clostridium acetobutylicum
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007686; AAK79737.1;
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
KW Complete proteome.
SQ SEQUENCE 489 AA: 54553 MW: 9002CB36AF1D3CD8 CRC64;

Query Match 51.8%; Score 59; DB 16; Length 489;
Best Local Similarity 57.9%; Pred. No. 0.26;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
OY 2 NSGDYKEDYGFHITRADM 20
DB 30 NDMNKETYGVSHITRNM 48

RESULT 3
ID 093290 PRELIMINARY: PRT: 487 AA.
AC 093290:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE AT4G24430/T2226.260.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Canlinci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY057715; AAL15345.1;
SQ SEQUENCE 487 AA: 55388 MW: 1684FED01571F262 CRC64;

Query Match 48.2%; Score 55; DB 10; Length 487;
Best Local Similarity 69.2%; Pred. No. 1.2;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
OY 5 DYKEDYGFHITR 17
DB 347 DYKDWEPFAHVTR 359

RESULT 4
ID 09STV1 PRELIMINARY: PRT: 646 AA.
AC 09STV1:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE LG27/30-like gene.
GN T22A6.260 OR AT4G24430.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Zimmermann W., Gruenisen A., Wambutt R., Bancroft I.,
RA Meves H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Zimmermann W., Gruenisen A., Wambutt R., Kalicki J., Wohlmann P.,
RA Smith A., Meves H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AL078637; CAB45078.1;
DR EMBL: AL161561; CAB79353.1;
SQ SEQUENCE 646 AA: 73608 MW: B5ACAD2021EAE0C3 CRC64;

Query Match 48.2%; Score 55; DB 10; Length 646;
Best Local Similarity 69.2%; Pred. No. 1.6;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
OY 5 DYKEDYGFHITR 17
DB 506 DYKDWEPFAHVTR 518

RESULT 5
ID 09P330 PRELIMINARY: PRT: 524 AA.
AC 09P330:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE 14 alpha-demethylase (14-a demethylase).
GN CYP51A1.
OS Venturia inaequalis.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
OC Dothideomycetes et Chaetothyrionmycetes incertae sedis; Venturiaceae;
OC Venturia.
OX NCBI_TaxID=5025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ENT27;
RA Schnabel G., Jones A.L.;
RT "Isolation and molecular characterization of the 14-a demethylase gene
RT from Venturia inaequalis and its polymorphic 5' flanking region."
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ENT54;
RA Schnabel G., Jones A.L.;
RT "Isolation and molecular characterization of the 14-a demethylase gene

RT of *Venturia inaequalis* and its polymorphic 5' flanking region."
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-F445;
RA Schabel G., Jones A.L.;
RT "Isolation and molecular characterization of the 14-a demethylase gene
from *Venturia inaequalis* and its polymorphic 5' flanking region."
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AF227920: AAF76464.1; -;
DR EMBL: AF262756: AAF71293.1; -;
DR EMBL: AF262757: AAF71294.1; -;
DR HSSP: P77901: 1E9X.
DR InterPro: IPR001128: Cytochrome_P450.
DR Pfam: PF00067: P450.1.
DR PRINTS: PR00385: P450.1.
DR PROSITE: PS00086: CYTOCHROME_P450; UNKNOWN_1.
DR Heme: Methyltransferase; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 524 AA: 59075 MW: AACFE2A1CD7A84B3 CRC64;

Query Match 45.6%; Score 52; DB 3; Length 524;
Best Local Similarity 56.2%; Pred. No. 4;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 NSGDYKEDYGFANTR 17
DB 435 NGGEKEDYGYGLIRK 450

RESULT 6
OY0A8 PRELIMINARY; PRT: 220 AA.
ID O9Y0A8:
AC O9Y0A8:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Type 2 adenylate kinase.
OS *Cryptosporidium parvum*.
OC *Cryptosporidium* parvum.
OC Eukaryota: Alveolata: Apicomplexa: Coccidia: Elmeriida:
OC *Cryptosporidiidae*: *Cryptosporidium*.
OX NCBI_TaxID=3807;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KSU-1;
RA Riordan C.E., Langreth S.G., Sanchez L.B., Upton S.J., Kayser O.,
RA Zhu G., Keithly J.S.;
RT "Preliminary Evidence for a Mitochondrion in *Cryptosporidium parvum*:
Phylogenetic and Therapeutic Implications."
RL J. Eukaryot. Microbiol. 0:0-0(1999).
CC -1- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
DR EMBL: AF163320: AAD45371.1; -;
DR HSSP: P07170: 1AKY.
DR InterPro: IPR000850: Adenylate_kin.
DR Pfam: PF00406: adenylatekinase.1.
DR PRINTS: PR00094: ADENYLATE_KINASE.
DR Prodom: PD000657: Adenylate_kin.1.
DR PROSITE: PS00113: ADENYLATE_KINASE.1.
KV Kinase; Transferase
SQ SEQUENCE 220 AA: 24228 MW: CA1D699A68A6DF CRC64;

Query Match 43.9%; Score 50; DB 5; Length 220;
Best Local Similarity 53.3%; Pred. No. 3.2;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 7 KEDIGFANITRADML 21
DB 27 KKEYGLAHSTGDM 41

RESULT 7
O9KRS8 PRELIMINARY; PRT: 478 AA.
ID O9KRS8

AC O9KRS8:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 6-phospho-beta-glucosidase.
GN VC1358.
OS *Vibrio cholerae*.
OC Bacteria: Proteobacteria: gamma subdivision: Vibrionaceae: *Vibrio*.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unsayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
cholerae".
RT Nature 406:477-483(2000).
RL Nature 406:477-483(2000).
DR EMBL: AE004233: AAF94712.1; -;
DR HSSP: P11546: 1PBG.
DR TIGR: VC1558; -;
DR InterPro: IPR001360: GH_1.
DR Pfam: PF00232: Glyco_hydro_1; 1.
DR PRINTS: PR00131: GLYDRLASE1.
DR Prodom: PD000650: GH_1; 1.
DR PROSITE: PS00572: GLYCOSYL_HYDROL_FL_1; UNKNOWN_1.
DR PROSITE: PS00653: GLYCOSYL_HYDROL_FL_2; 1.
KV Complete proteome.

SO SEQUENCE 478 AA: 55214 MW: DC7642B983300F87 CRC64;

Query Match 42.1%; Score 48; DB 16; Length 478;
Best Local Similarity 41.2%; Pred. No. 17;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 3 SGDYKEDYGFANTRAD 19
DB 434 TGEYKRRYGFIVDKD 450

RESULT 8
O9Z5E5 PRELIMINARY; PRT: 499 AA.
ID O9Z5E5:
AC O9Z5E5:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE levansucrase.
GN SABC.
OS *Paenibacillus polymyxa* (*Bacillus polymyxa*).
OC Bacteria: Firmicutes: *Bacillus*/Clostridium group: Bacillales;
OC *Bacillaceae*: *Paenibacillus*.
OX NCBI_TaxID=1406;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CF43;
RX MEDLINE=21042000; PubMed=11200435;
RA Beazate S., Aymerich S., Chamber R., Czarnez S., Berge O., Heulin T.;
RT "Disruption of the *Paenibacillus polymyxa* levansucrase gene impairs
its ability to aggregate soil in the wheat rhizosphere."
RL Environ. Microbiol. 2:335-342(2000).
DR EMBL: AJ133737: CAB39327.1; -;
DR InterPro: IPR003469: Glyco_hydro_68.
DR Pfam: PF02435: Glyco_hydro_68; 1.
SQ SEQUENCE 499 AA: 55317 MW: E646CD986292336C CRC64;

Query Match 42.1%; Score 48; DB 2; Length 499;
Best Local Similarity 50.0%; Pred. No. 17;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY 2 NSGPKEDYGFPHITRADML 21
 DB 33 DASDKTDNAFTQITRNDML 52

RESULT 9

08MYW3 PRELIMINARY: PRT: 582 AA.
 AC 08MYW3.
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DE Hypothetical 66.3 kDa protein.
 GN Pp13624.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhang P.P., Zhou X.M., Jiang H.Q., Huang Y., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF318335; AAL55842.1;
 KW Hypothetical protein.
 SQ SEQUENCE 582 AA; 66322 MW; B847F093D01B1982 CRC64;

Query Match 42.1%; Score 48; DB 4; Length 582;
 Best Local Similarity 31.2%; Pred. No. 21;
 Matches 10; Conservative 4; Mismatches 2; Indels 16; Gaps 1;

OY 5 DYKEDYG-----FAHITRADM 20
 DB 297 DYEEKYGRMPREWCMAERIAVEFCHVTRAEI 328

RESULT 10

09BRR2 PRELIMINARY: PRT: 670 AA.
 AC 09BRR2.
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Similar to RIKEN cDNA 3100002B05 gene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tissue-PLACENTA;
 RC Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC006116; AAH06116.1;
 SQ SEQUENCE 670 AA; 76351 MW; 8057AAC082136EC7 CRC64;

Query Match 42.1%; Score 48; DB 4; Length 670;
 Best Local Similarity 31.2%; Pred. No. 24;
 Matches 10; Conservative 4; Mismatches 2; Indels 16; Gaps 1;

OY 5 DYKEDYG-----FAHITRADM 20
 DB 268 DYEEKYGRMPREWCMAERIAVEFCHVTRAEI 299

RESULT 11

09BYO2 PRELIMINARY: PRT: 673 AA.
 AC 09BYO2.
 DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Hypothetical 76.7 kDa protein.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhao X., Wan D., Han L., Gu J.;
 RT "cDNA clone C63R on chromosome 17p13.3."
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF246287; AAK27973.1;
 KW Hypothetical protein.
 SQ SEQUENCE 673 AA; 76667 MW; 8257C0D59E9385C1 CRC64;

Query Match 42.1%; Score 48; DB 4; Length 673;
 Best Local Similarity 31.2%; Pred. No. 24;
 Matches 10; Conservative 4; Mismatches 2; Indels 16; Gaps 1;

OY 5 DYKEDYG-----FAHITRADM 20
 DB 297 DYEEKYGRMPREWCMAERIAVEFCHVTRAEI 328

RESULT 12

09CX82 PRELIMINARY: PRT: 832 AA.
 AC 09CX82.
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 3100002B05RIK protein.
 GN 3100002B05RIK
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRIONIC HEAD;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Aachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann W., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
 RA Suzuki H., Toyoko-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK019408; BAB31707.1;
 DR MGD: MGI:1915549; 3100002B05RIK.
 SQ SEQUENCE 832 AA; 94464 MW; C22B4496491ACDC CRC64;

Query Match 42.1%; Score 48; DB 11; Length 832;
 Best Local Similarity 31.2%; Pred. No. 31;
 Matches 10; Conservative 4; Mismatches 2; Indels 16; Gaps 1;

OY 5 DYKEDYG-----FAHITRADM 20
 DB 297 DYEEKYGRMPREWCMAERIAVEFCHVTRAEI 328

RESULT 13
09SZJ8 PRELIMINARY; PRT: 289 AA.
ID 09SZJ8
AC 09SZJ8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Hypothetical 33.4 kDa protein.
GN F20D10.80 OR ATAG37960.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Beyer M., Wedler H., Kutzner M., Wambutt R., Bancroft I., Newes H.W.,
Mayer K.F.X., Schueller C.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Wedler H., Kutzner M., Wambutt R., Mewes H.W., Lemcke K.,
Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL035538; CAB37536.1; -
DR EMBL: AL161592; CAB80461.1; -
KW Hypothetical protein.
SQ SEQUENCE 289 AA; 3385 MW; C0BFEE6AF7EEF554 CRC64;
SO
Query Match 41.2%; Score 47; DB 10; Length 289;
Best Local Similarity 53.8%; Pred. No. 14;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 5 DYKEDYGFHITR 17
Db 157 DYKDMFYAHVNR 169
RESULT 14
06S620 PRELIMINARY; PRT: 435 AA.
AC 06S620;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MST1 protein (Fragment).
GN MYST1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=COLUMBIA, AND CV. CO;
RX MEDLINE=20267440; PubMed=10809443;
RA Tavares R., Aubourg S., Decharny A., Kreis M.;
RT "Organization and structural evolution of four multigene families in
Arabidopsis thaliana: ATLCA, ATLGT, ATWST and ATHD-GL2.";
RL Plant Mol. Biol. 42:703-717(2000).
DR EMBL: Y16848; CAA76417.1; -
FT NON_TER 1
SQ SEQUENCE 435 AA; 49882 MW; 9591F93B67214562 CRC64;

Query Match 41.2%; Score 47; DB 10; Length 435;
Best Local Similarity 53.8%; Pred. No. 22;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 5 DYKEDYGFHITR 17
Db 303 DYKDMFYAHVNR 315
RESULT 15
098R03 PRELIMINARY; PRT: 567 AA.
ID 098R03
AC 098R03;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE LIPOPROTEIN.
GN MYPU_2070.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Hellig R., Ferris S., Barbe V., Samson D., Galisson F.,
Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
Mycoplasma pulmonis.";
RL Nucleic Acids Res. 29:2145-2153(2001).
DR EMBL: AL445563; CAC13380.1; -
DR Mypulast; MYPU_2070; -
DR InterPro: IPR002057; Isopen_N_synth.
DR PROSITE: PS00186; IPNS_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 567 AA; 64792 MW; C6D953B917B76574 CRC64;
SO
Query Match 41.2%; Score 47; DB 16; Length 567;
Best Local Similarity 44.4%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
QY 2 NSGDYKEDYGFHITRAD 19
Db 435 HGNKVDYVFTHTNRTN 452

Search completed: June 16, 2003, 12:09:34
Job time : 7.29303 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:56:18 ; Search time 1.33402 Seconds
(without alignments)
652.918 Million cell updates/sec

```
Title: US-09-986-682B-1
Perfect score: 114
Sequence: 1 MNSGDYKEDYGFAHITRADML 21
```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	ID	Description
1	68	59.6	473	1	SACB_BACST
2	68	59.6	473	1	SACB_BACST
3	66	57.9	472	1	SACB_BACAM
4	48	42.1	295	1	MURB_RICCN
5	48	42.1	295	1	MURB_RICPR
6	46	40.4	217	1	KAD_BACSU
7	44	38.6	342	1	CCR6_HUMAN
8	44	38.6	533	1	RRPB_CVPEFS
9	44	38.6	541	1	NASY_MYXXA
10	43	37.7	212	1	KAD_STRPY
11	43	37.7	446	1	YCAJ_COXBU
12	43	37.7	958	1	AMYG_DEBOC
13	43	37.7	1094	1	EX5B_MYCTU
14	43	37.7	1297	1	PUR4_HAETN
15	42.5	37.3	587	1	CO8B_ONCMY
16	42	36.8	125	1	RNPA_CLOPE
17	42	36.8	193	1	KAD_CYPCA
18	42	36.8	202	1	UMPK_ARATH
19	42	36.8	204	1	UMPK_YEAST
20	42	36.8	214	1	KAD_MTCGA
21	42	36.8	220	1	KAD1_SCHPO
22	42	36.8	479	1	BGLA_ECOLI
23	42	36.8	580	1	MP12_HUMAN
24	42	36.8	600	1	KU70_RH1AP
25	41.5	36.4	343	1	GUNC_CLOSF
26	41.5	36.4	343	1	GUNC_CLOTM
27	41	36.0	198	1	KAD5_HUMAN
28	41	36.0	231	1	RAD5_PRUAR
29	41	36.0	380	1	COTB_BACSU
30	41	36.0	561	1	CP51_USTWA
31	41	36.0	986	1	GUN2_CLOS
32	40.5	35.5	504	1	TABL_HUMAN
33	40	35.1	130	1	LYCK_SHEEP
					P05468 bacillus st
					P06565 bacillus su
					P21130 bacillus am
					O991t8 rickettsia
					O993s7 rickettsia
					P16304 bacillus su
					O00574 homo sapien
					P18457 porcine tra
					P95329 MYXOCOCC
					P82549 streptococ
					P33918 coxiella bu
					P22861 debaryomyce
					P96590 mycobacteri
					P43847 haemophilus
					O98x85 oncorhynch
					O8h26 clostridium
					P1115 cyprinus ca
					O04905 arabidopsis
					P15700 saccharomyc
					O5352 mycoplasma
					P33075 schizosacch
					O46829 escherichia
					P30305 homo sapien
					P26228 rhipicephal
					O23340 clostridium
					P07985 clostridium
					O9y6x8 homo sapien
					O24464 prunus arme
					P07789 bacillus su
					P46602 ussiliago ma
					P23659 clostridium
					O15750 homo sapien
					P80190 ovis aries

34	40	35.1	279	1	HEM6_RICPR	092626	ricketsia
35	40	35.1	295	1	YVC1_BACSU	006973	bacillus su
36	40	35.1	366	1	A2HS_CAVPO	070159	cavia porce
37	40	35.1	374	1	G3PA_CHURE	P50362	chlamydomon
38	40	35.1	465	1	YHUA_ECOLI	P37157	escherichia
39	40	35.1	504	1	RA52_YEAST	P67778	saccharomy
40	40	35.1	617	1	ASMA_ECOLI	P28493	escherichia
41	40	35.1	920	1	VLGB_HSVSM	004464	herpesviral
42	40	35.1	1297	1	PUR4_PASMU	094464	pasturella
43	40	35.1	3415	1	POIG_POWVL	004558	t genome po
44	39.5	34.6	340	1	ARQC_AOUAE	0057724	agutifer aeo
45	39.5	34.6	933	1	VLGB_HSVAI	004463	herpesviral

ALIGNMENTS

```

RESULT 1
SACB_BACST STANDARD: PRT: 473 AA.
AC P94468: 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE levensucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
transferase) (Sucrose 6-fructosyl transferase).
GN SACH OR SURB.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12980;
RX MEDLINE=98007870; PubMed=9349714;
RA Li Y., Tricas J.A., Ferenci T.,
RT "A novel levensucrase-levanase gene cluster in Bacillus
RL stearothermophilus ATCC12980".
CC Biochim. Biophys. Acta 1353:203-208(1997).
CC -I CATALYTIC ACTIVITY: Sucrose + [(2,6)-beta-D-fructosyl](n) =
CC glucose + [(2,6)-beta-D-fructosyl](n+1).
CC -I SUBCELLULAR LOCATION: Secreted.
CC -I SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on way
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC CC
CC EMBL: U34874; AAB97111.1; -.
CC DR InterPro: IPR003469; Glyco_hydro_68.
CC DR Pfam: PF02435; Glyco_hydro_68; 1.
CC KW Transferase; Glycosyltransferase; Signal.
CC FT SIGNAL 1 29 BY SIMILARITY.
CC FT CHAIN 30 473 LEVANSUCRASE.
CC SO SEQUENCE 473 AA; 52854 MW; 12E7A272EC2B71E CRC64;
-----
QY Query Match 59.6%; Score 68; DB 1; Length 473;
Best Local Similarity 65.0%; Pred. NO. 0.0023;
Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0.
DB 2 NSCDKEDYGFANHTRDML 21
1 ||| ||| ||| |||
33 NOKPYKERYGISHTRDML 52

```

DT 01-NOV-1988 (rel. 09, last sequence update)
 DT 15-JUN-2002 (rel. 41, last annotation update)
 DE Levanisucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
 DE transferase) (Sucrose 6-fructosyl transferase).
 GN SACH.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168.
 RA MEDLINE=85295507; PubMed=2993818;
 RA Steinmetz M., Le Coq D., Aymerich S., Gonzy-Treboul G., Gay P.;
 RT "The DNA sequence of the gene for the secreted Bacillus subtilis
 RT enzyme levanisucrase and its genetic control sites."
 RL Mol. Gen. Genet. 200;220-228(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Denizot F.;
 RL Submitted (APR-1997) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borliss R., Bourcier L., Brans A., Braun M., Briganti S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Erington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Mestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
 RA Prescan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
 RA Sorokin A., Tacconi E., Serror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni K.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
 RA Viari A., Wambuit R., Wedler E., Wedler H., Weltzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis."
 RL Nature 390:249-256(1997).
 RN [4]
 RP SEQUENCE OF 1-62 FROM N.A.
 RC STRAIN=168 / PY79;
 RA MEDLINE=84178454; PubMed=6424671;
 RA Fouet A., Arnaud M., Klier A., Rapoport G.;
 RT "Characterization of the precursor form of the exocellular
 RT levanisucrase from Bacillus subtilis."
 RL Biochem. Biophys. Res. Commun. 119:795-800(1984).
 RN [5]
 RP SEQUENCE OF 1-68 FROM N.A.
 RC MEDLINE=87008406; PubMed=2428811;
 RA Shimotsu H., Henner D.J.;
 RT "Modulation of Bacillus subtilis levanisucrase gene expression by
 RT sucrose and regulation of the steady-state mRNA level by sacY and
 RT sacZ genes."
 RL J. Bacteriol. 168:380-388(1986).
 CC -i- CATALYTIC ACTIVITY: Sucrose + [(2,6)-beta-D-fructosyl](N) -

CC glucose + [(2,6)-beta-D-fructosyl](N+1).
 CC -i- SUBCELLULAR LOCATION: Secreted.
 CC -i- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M14202; AAA22725.1; -
 CC EMBL: 294043; CAB08015.1; -
 CC EMBL: 299121; CAB15450.1; -
 CC EMBL: X01987; AAA22724.1; -
 CC EMBL: X02730; CAA26513.1; -
 CC PIR: A25040; A25040.
 CC PIR: S07309; S07309.
 CC Subtilisin; BG10388; sacB.
 CC InterPro: IPR003469; Glyco_hydro_68.
 CC Pfam: PF02435; Glyco_hydro_68; 1.
 CC Transferrase; Glycosyltransferase; Signal; Complete proteome.
 CC FT SIGNAL 1 29
 CC CHAIN 30 473 LEVANSUCRASE.
 CC CONFLICT 12 12 V -> I (IN REF. 4).
 CC SEQUENCE 473 AA: 52971 MW: 38FP2F571B41D5B0 CRC64:
 CC
 CC Query Match 59.6%; Score 68; DB 1; Length 473;
 CC Best Local Similarity 65.0%; Pred. No. 0.0023;
 CC Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 CC
 CC OY 2 NSGDYKEDYGFAPTRADL 21
 CC DB 33 NQPKYKRYGSHITRDM 52
 CC
 CC RESULT 3
 CC SACH_BACAM STANDARD; PRT; 472 AA.
 CC ID P21130.
 CC AC 01-FEB-1991 (rel. 17, Created)
 CC DT 01-FEB-1991 (rel. 17, Last sequence update)
 CC DT 15-DEC-1998 (rel. 37, Last annotation update)
 CC DE Levanisucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
 CC DE transferase) (Sucrose 6-fructosyl transferase).
 CC GN SACH.
 OS Bacillus amyloliquefaciens.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1390;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 23844;
 RA MEDLINE=91092506; PubMed=2265762;
 RA Tang L.B., Lenstra R., Borchert T.V., Vasantha N.;
 RT "Isolation and characterization of levanisucrase-encoding gene from
 RT Bacillus amyloliquefaciens."
 RL Gene 96:89-93(1990).
 CC -i- CATALYTIC ACTIVITY: Sucrose + [(2,6)-beta-D-fructosyl](N) -
 CC glucose + [(2,6)-beta-D-fructosyl](N+1).
 CC -i- SUBCELLULAR LOCATION: Secreted.
 CC -i- INDUCTION: BY SUCROSE.
 CC -i- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X52988; CAA37179.1; -

```

DR PIR: J00802; J00802.
DR PIR: S11739; S11739.
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 29
FT CHAIN 30 472
SQ SEQUENCE 472 AA; 52859 MM; F38592D272677E7D CRC64;

Query Match 57.9%; Score 66; DB 1; Length 472;
Best Local Similarity 65.0%; Pred. No. 0.0048;
Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 2 NSGDYKEDYGFHITRADML 21
Db 33 NQAKERYGVSHTRHML 52

RESULT 4
MURB_RICCN STANDARD: PRT; 295 AA.
AC Q92IT8:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-N-acetylenolpyruvoylglycosamine reductase (EC 1.1.1.158) (UDP-N-
  acetylmuramate dehydrogenase).
GN MURB OR RC0332.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiase; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Rensio-Audiffren P., Fournier P.-E., Barbe V.,
RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
CC -1- FUNCTION: Cell wall formation (By similarity).
CC -1- CATALYTIC ACTIVITY: UDP-N-acetylmuramate + NADP(+) = UDP-N-
  acetyl-3-O-(1-carboxyvinyl)-D-glucosamine + NADPH.
CC -1- COFACTOR: FAD (By similarity).
CC -1- PATHWAY: Peptidoglycan biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE MURB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE008598; AAI02870.1; ALT_INIT.
DR InterPro: IPR003170; MURB.
DR InterPro: IPR001575; Oxiid_FAD_bind.
DR Pfam: PF01565; FAD_binding_4; 1.
DR Pfam: PF02873; MURB_C; 1.
KW Oxidoreductase; NADP; Flavoprotein; FAD; Cell wall; Cell division;
KW Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 295 AA; 32695 MM; 43EFA191E71F3678 CRC64;

Query Match 42.1%; Score 48; DB 1; Length 295;
Best Local Similarity 53.8%; Pred. No. 2.2;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 4 GDYKEDYGFHIT 16
Db 9 GEYKKDYNLKHUT 21

```

```

RESULT 5
MURB_RICPR STANDARD: PRT; 295 AA.
AC Q92DS7:
DT 30-MAY-2000 (Rel. 39, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-N-acetylenolpyruvoylglycosamine reductase (EC 1.1.1.158) (UDP-N-
  acetylmuramate dehydrogenase).
GN MURB OR RP248.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiase; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RT "The genome sequence of Rickettsia prowazekii and the origin of
  mitochondria.";
RL Nature 396:133-140(1998).
CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: UDP-N-acetylmuramate + NADP(+) = UDP-N-
  acetyl-3-O-(1-carboxyvinyl)-D-glucosamine + NADPH.
CC -1- COFACTOR: FAD (BY SIMILARITY).
CC -1- PATHWAY: Peptidoglycan biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE MURB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ235271; CAA14710.1; ALT_INIT.
DR InterPro: IPR003170; MURB.
DR InterPro: IPR001575; Oxiid_FAD_bind.
DR Pfam: PF01565; FAD_binding_4; 1.
DR Pfam: PF02873; MURB_C; 1.
KW Peptidoglycan synthesis; Cell wall; Cell division; Oxidoreductase;
KW NADP; Flavoprotein; FAD; Complete proteome.
SQ SEQUENCE 295 AA; 32751 MM; D59581825E9F0BAC CRC64;

Query Match 42.1%; Score 48; DB 1; Length 295;
Best Local Similarity 53.8%; Pred. No. 2.2;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 4 GDYKEDYGFHIT 16
Db 9 GEYKKDYNLKHUT 21

RESULT 6
KAD_BACSU STANDARD: PRT; 217 AA.
AC P16304:
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adenylate kinase (EC 2.7.4.3) (ATP-AMP transphosphorylase)
DE (Superoxide-inducible protein 16) (Sot16).
GN ADK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;

```

RN [1] SEQUENCE FROM N.A.
 RP MEDLINE=90292990; PubMed=2113521;
 RA Nakamura K., Nakamura A., Takamatsu H., Yoshikawa H., Yamane K.;
 RT "Cloning and characterization of a Bacillus subtilis gene homologous
 RT to E. coli secY.";
 RL J. Biochem. 107:603-607(1990).
 RN [2] SEQUENCE FROM N.A.
 RP STRAIN=168 / Marburg;
 RX MEDLINE=96186897; PubMed=8635744;
 RA Sun J.W., Boylan S.A., Oh S.H., Price C.W.;
 RT "Genetic and transcriptional organization of the Bacillus subtilis
 RT spc-alpha-region.";
 RL Gene 169:17-23(1996).
 RN [3] SEQUENCE OF 1-116 FROM N.A.
 RP MEDLINE=90221911; PubMed=2139212;
 RA Yoshikawa H., Doi R.H.;
 RT "Sequence of the Bacillus subtilis spectinomycin resistance gene
 RT region.";
 RL Nucleic Acids Res. 18:1647-1647(1990).
 RN [4] SEQUENCE OF 1-99 FROM N.A.
 RP STRAIN=168;
 RX MEDLINE=90251170; PubMed=2110998;
 RA Suh J.W., Boylan S.A., Thomas S.M., Dolan K.M., Oliver D.B.,
 RA Price C.W.;
 RT "Isolation of a secY homologue from Bacillus subtilis: evidence for a
 RT common protein export pathway in eubacteria.";
 RL Mol. Microbiol. 4:305-314(1990).
 RN [5] SEQUENCE OF 1-20.
 RP STRAIN=168 / IS58;
 RX MEDLINE=97443988; PubMed=9298659;
 RA Anselmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
 RA Hecker M.;
 RT "First steps from a two-dimensional protein index towards a response-
 RT regulation map for Bacillus subtilis.";
 RL Electrophoresis 18:1451-1463(1997).
 CC -!- FUNCTION: THIS SMALL UBQUITOUS ENZYME IS ESSENTIAL FOR
 CC MAINTENANCE AND CELL GROWTH.
 CC -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
 CC -!- SUBUNIT: MONOMER.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- INDUCTION: BY SUPEROXIDE.
 CC -!- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 CC EMBL: D00619; BAA00496.1; -
 CC EMBL: L47971; AAB06820.1; -
 CC EMBL: M31102; AAB59119.1; -
 CC EMBL: X51329; CAA35713.1; -
 CC EMBL: Z99104; CAB11913.1; -
 CC PIR: J50492; J50492.
 CC PIR: S08630; S08630.
 CC PIR: S12684; S12684.
 CC HSSP: P27142; 1ZIN.
 CC Subtilist: BG10446; adk.
 CC InterPro: IPR000850; Adenylate_kin.
 CC Pfam: PF00406; adenylate_kinase.1.
 CC PRINTS: PR00094; ADENYLTKINASE.
 CC ProDom: PD000657; Adenylate_kin.1.
 CC PROSITE: PS00113; ADENYLATE_KINASE.1.
 CC Transferrase; kinase; ATP-binding; Complete proteome.
 FT NP_BIND 7 15 ATP (BY SIMILARITY).

SQ SEQUENCE 217 AA; 24119 MW; ECB9ECFAF26A1E90 CRC64;
 Query Match 40.4%; Score 46; DB 1; Length 217;
 Best Local Similarity 61.5%; Pred. No. 3.3;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 8 EDYGFHITTRADM 20
 1111 11: 11
 DB 22 EDYGIPHISTGDM 34
 RESULT 7
 CCR6_HUMAN
 ID CCR6_HUMAN STANDARD; PRT; 342 AA.
 AC 000574; 000575;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE C-X-C chemokine receptor type 6 (CXCR-6) (G protein-coupled
 DE receptor bonzo) (G protein-coupled receptor STRL33).
 GN CXCR6 OR BONZO OR STRL33 OR TYMSR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1] SEQUENCE FROM N.A.
 RP MEDLINE=97373958; PubMed=9230441;
 RX Deng H.K., Unutmaz D., Kewellramani V.N., Littman D.R.;
 RT "Expression cloning of new receptors used by simian and human
 RT immunodeficiency viruses.";
 RL Nature 388:296-300(1997).
 RN [2] SEQUENCE FROM N.A.
 RP MEDLINE=97311099; PubMed=9166430;
 RX Liao F., Alkhalib G., Peden K.W.C., Sharma G., Berger E.A.,
 RA Farber J.M.;
 RT "STRL33, A novel chemokine receptor-like protein, functions as a
 RT fusion cofactor for both macrophage-tropic and T cell line-tropic
 RT HIV-1.";
 RL J. Exp. Med. 185:2015-2023(1997).
 RN [3] SEQUENCE FROM N.A.
 RP TISSUE=Blood;
 RC TISSUE=Blood;
 RX MEDLINE=974431687; PubMed=9285716;
 RA Loeschner M., Amara A., Oberlin E., Brass N., Legler D.F.,
 RA Loeschner P., D'Apuzzo M., Meese E.U., Rousset D., Virelizier J.L.,
 RA Baggiolini M., Arenzana-Seisdedos F., Moser B.;
 RT "TYMSR, a putative chemokine receptor selectively expressed in
 RT activated T cells, exhibits HIV-1 coreceptor function.";
 RL Curr. Biol. 7:652-660(1997).
 CC -!- FUNCTION: RECEPTOR FOR THE C-X-C CHEMOKINE CXCL16. USED AS A
 CC CORECEPTOR BY SIVS AND BY STRAINS OF HIV-2 AND M-TROPIC HIV-1.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN LYMPHOID TISSUES AND ACTIVATED T
 CC CELLS.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 CC EMBL: AF007545; AAB64221.1; -
 CC EMBL: U73529; AAB61456.1; -
 CC EMBL: U73531; AAB61457.1; -
 CC EMBL: Y13248; CAA73698.1; -
 CC MIM: 605163; -
 CC InterPro: IPR000276; GPCR_Rhodpsn.
 CC Pfam: PF00001; 7tm_1; 1.

```

DR PRINTS: PR00237; GPCRHDOPSN.
DR PROSITE: PS00237; G-PROTEIN RECEPTOR FL 1; 1.
DR PROSITE: PS00262; G-PROTEIN RECEPTOR FL 2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Polymorphism.
FT DOMAIN 1 32
FT TRANSMEM 33 59
FT DOMAIN 60 69
FT TRANSMEM 69 88
FT DOMAIN 90 103
FT TRANSMEM 104 125
FT DOMAIN 126 143
FT TRANSMEM 144 164
FT TRANSMEM 165 187
FT TRANSMEM 188 215
FT TRANSMEM 216 231
FT TRANSMEM 232 259
FT DOMAIN 260 275
FT TRANSMEM 276 293
FT DOMAIN 294 342
FT DISULFID 102 180
FT CARBOHYD 16 16
FT VARIANT 25 25
D->A (IN STRU3.3).
/FTID=VAR_003506.
SQ SEQUENCE 342 AA; 39280 MW; 9FBC025556D1082E CRC64;

Query Match 38.6%; Score 44; DB 1; Length 342;
Best Local Similarity 61.5%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGEFA 13
   1 1 1 1 1 1 1
Db 1 MAEHYHEDYGEFS 13

RESULT 8
RRPB_CVFPs STANDARD: PRT: 533 AA.
ID RRPB_CVFPs STANDARD: PRT: 533 AA.
AC P18457;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA-directed RNA polymerase (ORF1B) (EC 2.7.7.48) (Fragment).
OS Porcine transmissible gastroenteritis coronavirus (strain F8772/70).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11150;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91188698; PubMed=1964522;
RA Britton P., Page K.W.;
RT "Sequence of the S gene from a virulent British field isolate of
RL transmissible gastroenteritis virus.";
RT Virus Res. 18:71-80(1990).
CC -1- FUNCTION: THE RNA DEPENDENT RNA POLYMERASE OF CORONAVIRUSES IS
CC A MULTIFUNCTIONAL PROTEIN: IT CONTAINS THE ACTIVITIES NECESSARY
CC FOR THE TRANSCRIPTION OF NEGATIVE STRANDED RNA, LEADER RNA,
CC SUBGENOMIC MRNAS AND PROGENY VIRION RNA.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(N)
CC -1- MISCELLANEOUS: THIS PROTEIN IS EXPRESSED BY AN EFFICIENT RIBOSOMAL
CC FRAMESHIFTING MECHANISM. RIBOSOMAL FRAMESHIFTING IS AN ELEGANT
CC MECHANISM FOR REGULATING THE SYNTHESIS OF SEVERAL PROTEINS IN A
CC WELL BALANCED MANNER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: X53128; CAA37284.1; -

```

```

DR PIR: A43489; A43489.
KW Transferrase; RNA-directed RNA polymerase; Helicase.
FT NON_TER 1
SQ SEQUENCE 533 AA; 60270 MW; B1AE8C49DBA7FC5E CRC64;

Query Match 38.6%; Score 44; DB 1; Length 533;
Best Local Similarity 53.8%; Pred. No. 18;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 8 EDYGFHATRADM 20
   1 1 1 1 1 1 1
Db 114 EDYGFHVFQDV 126

RESULT 9
MASY_MYXXA STANDARD: PRT: 541 AA.
ID MASY_MYXXA STANDARD: PRT: 541 AA.
AC P95329;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Malate synthase (EC 4.1.3.2).
GN MLS.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cytophacteriineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DK1622;
RA Salmi D., Creighton C., Youderian P.;
RT "Myxococcus xanthus fumarate hydratase, major proteosome subunit, and
RL acyl-CoA oxidase genes.";
CC Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-malate + CoA = acetyl-CoA + H(2)O +
CC glyoxylate.
CC -1- PATHWAY: SECOND STEP IN GLYOXYLATE BYPASS, AN ALTERNATIVE TO THE
CC TRICARBOXYLIC ACID CYCLE (IN BACTERIA, FUNGI AND PLANTS).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: BELONGS TO THE MALATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF013216; AAB97829.1; -
DR InterPro: IPR001465; Malate_synthase.
DR Pfam: PF01274; Malate_synthase.1.
DR PROSITE: PS00510; MALATE_SYNTHASE; 1.
KW Lyase; Glyoxylate bypass; Tricarboxylic acid cycle.
SQ SEQUENCE 541 AA; 60524 MW; DAD34265955B6E26 CRC64;

Query Match 38.6%; Score 44; DB 1; Length 541;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 7 KEDYGFHATRADM 21
   1 1 1 1 1 1 1
Db 501 KERYGAHLERARV 515

RESULT 10
KAD_STRPY STANDARD: PRT: 212 AA.
ID KAD_STRPY STANDARD: PRT: 212 AA.
AC P82549;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adenylate kinase (EC 2.7.4.3) (ATP-AMP transphosphorylase).

```

```

GN ADR OR SPY0074.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., Moshan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezeau S., Suvoirov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RN [2]
RP SEQUENCE OF 1-13 AND 207-212, AND MASS SPECTROMETRY.
RC STRAIN-JR34 / Serotype M6;
RA Hogan D.A., Du P., Stevenson T.I., Whilton M., Kilby G.W., Rogers J.,
RA Vanbogaelen R.A.;
RT "Two-dimensional gel electrophoresis map of Streptococcus pyogenes
RT proteins";
RL Submitted (MAY-2000) to the SWISS-PROT data bank.
CC -1- FUNCTION: THIS SMALL UBQUITOUS ENZYME IS ESSENTIAL FOR
CC MAINTENANCE AND CELL GROWTH.
CC -1- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- MASS SPECTROMETRY: MW=2369.98; METHOD=Electrospray.
CC -1- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE006478; AAK33203.1; -
DR HSSP: P27142; 1ZTN.
DR InterPro: IPR000850; Adenylate_kin.
DR Pfam: PF00406; adenylatekinase.1.
DR PRINTS: PR00094; ADENYLTKINASE.
DR ProDom: PD000657; Adenylate_kin; 1.
DR PROSITE: PS00113; ADENYLATE_KINASE; 1.
KW Transferase; Kinase; ATP-binding; Complete proteome.
FT NP_BIND 7 15
FT SEQUENCE 212 AA: 23700 MW: F98AA65E98B39DEE CRC64;
SQ

```

Query Match 37.7%; Score 43; DB 1; Length 212;
 Best Local Similarity 53.8%; Pred. No. 9.8;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

OY 8 EDYGFAMITRADM 20
   1 1 1 1 1 1 1
Db 22 EEFGLAHISTGDM 34

```

RESULT 11

```

YCAJ_COXBU STANDARD: PRT: 446 AA.
ID YCAJ_COXBU
AC P39918;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 49.9 kDa protein in SpoIIIE-ers intergenic region.
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Coxiella group; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Oswald W.;
RL Thesis (1994), Justus Liebig University / Frankfurt, Germany.
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Balroch A.;
RL Unpublished observations (AUG-1995).
CC -1- SIMILARITY: STRONG, TO E.COLI YCAJ AND H.INFLUENZAE HOMOLOG.
CC -1- SIMILARITY: TO YEAST YNL218W.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION: A FRAMESHIFT WAS
CC INTRODUCED IN POSITION 66 TO PRODUCE THIS ORF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X75627; CAA53291.1; ALT_FRAME.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003959; AAA_ATPase-centr.
DR InterPro: IPR000862; RRCdomain.
DR Pfam: PF00004; AAA; 1.
DR SMART: SM00382; AAA; 1.
KW Hypothetical protein; ATP-binding.
FT NP_BIND 51 58
FT SEQUENCE 446 AA: 49875 MW: D24877B3C6A4E04B CRC64;
SQ

```

Query Match 37.7%; Score 43; DB 1; Length 446;
 Best Local Similarity 57.1%; Pred. No. 21;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

OY 1 MNSGDYKEDYGFAN 14
   1 1 1 1 1 1 1
Db 377 MKSLDYGRDYVAH 390

```

RESULT 12

```

AMYG_DEBOC STANDARD: PRT: 958 AA.
ID AMYG_DEBOC
AC P22861; G92336;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Glucoamylase 1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GAM1.
OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX NCBI_TaxID=27300;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-ATCC 26076;
RX MEDLINE=91071592; PubMed=1979298;
RA Dohmen R.J., Strasser A.W.M., Dahlems U.M., Hollenberg C.P.;
RT "Cloning of the Schwanniomyces occidentalis glucoamylase gene (GAM1)
RT and its expression in Saccharomyces cerevisiae.";
RL Gene 95:111-121(1990).
RN [2]
RP SIMILARITY TO OTHER FAMILY 31 MEMBERS.
RX MEDLINE=92077121; PubMed=1743281;
RA Naim H.Y., Niemann T., Kleinhans U., Hollenberg C.P.,
RA Strasser A.W.M.;
RT Striking structural and functional similarities suggest that
RT intestinal sucrose-isomaltase, human lysosomal alpha-glucosidase and
RT Schwanniomyces occidentalis glucoamylase are derived from a common
RT ancestral gene.";
RL FEBS Lett 294:109-112(1991).
CC -1- FUNCTION: THIS GLUCOAMYLASE HAS A SPECIFICITY TOWARD BOTH
CC ALPHA-1,4 AND ALPHA-1,6 LINKAGES.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-

```


RT complete genome sequence."
RL Nature 393:537-544(1998).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Feldman J., Khouri H., Gill J., Mikula A.,
RA Bishal W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR: IT CATALYZES THE
CC UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
CC STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP
CC (BY SIMILARITY).
CC -I- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
CC ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
CC phosphooligonucleotides.
CC -I- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD
CC (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVRD SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z92772; CAB07119.1; -
DR EMBL: AE006961; AAK44882.1; -
DR HSSP: P56255; 2PJR.
DR TIGR: MT0658; -
DR TuberculList; RV0630C; -
DR InterPro: IPR004586; RECB.
DR InterPro: IPR000212; UVRD-helicase.
DR Pfam: PF00580; UVRD-helicase; 1.
DR TIGRFAMS: TIGR00609; recb; 1.
KW Hydroxylase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
KW DNA repair; Complete proteome.
FT NP_BIND 21 ATP (POTENTIAL).
FT FT 28
SQ SEQUENCE 1094 AA; 118721 MW; 31262D376875C201 CRC64;

Query Match 37.7%; Score 43; DB 1; Length 1094;
Best Local Similarity 53.3%; Pred. No. 54;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 GDYKEDYGFAHITRA 18
Db 996 GDRADYGFERLTER 1010

RESULT 14
PUR4_HAEIN
ID PUR4_HAEIN STANDARD: PRT: 1297 AA.
AC P43847;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphoribosylformylglycinamide synthase (EC 6.3.5.3) (FGAM
DE synthase) (Formylglycinamide ribotide amidotransferase) (FGARAT).
DE PURL OR H10752.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;

```

RX MEDLINE=95350630; PubMed=7542800;
RA Flischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Cocayne J.D.,
RA Scott J.D., Shilley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-
CC ribosyl)glycanamide + L-glutamine + H(2)O = ADP + phosphate + 2-
CC (formamido)-N(1)-(5-phospho-D-ribo-syl)acetamide + L-glutamate.
CC -1- PATHWAY: De novo purine biosynthesis; fourth step.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE FGAMS
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32759; AAC22411.1; ALT_INIT.
CC TIGR: HI0752.
CC InterPro: IPR000728; AIRS_related.
CC Pfam: PF00586; AIRS; 1.
CC Pfam: PF02769; AIRS_C; 2.
CC Putative biosynthesis; Ligase: ATP-binding; Glutamine amidotransferase;
CC Complete proteome.
CC NP_BIND 308 319 ATP (POTENTIAL).
CC ACET_SITE 1138 1138 GATASE (BY SIMILARITY).
CC SO SEQUENCE. 1297 AA; 142747 MW; DCD23A9DA666E93 CRC64;
OY 1 MNSGDYKEDYGEFAHTR 17
Db 457 WDSGSKEDLPASVOR 473

RESULT 15
COBB_ONCMY STANDARD: PRT: 587 AA.
AC Q90X85;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement component C8 beta chain precursor.
GN C8B.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBL_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RA Zarkadis I.K., Kazantzis A., Sfyroera G., Lambiris J.D.;
RT "Cloning and characterization of eighth beta component of complement
RT in rainbow trout."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: C8 IS A CONSTITUENT OF THE MEMBRANE ATTACK COMPLEX. C8
CC BINDS TO THE C5B-7 COMPLEX, FORMING THE C5B-8 COMPLEX. C5-B8 BINDS

```

```

CC C9 AND ACTS AS A CATALYST IN THE POLYMERIZATION OF C9 (BY
CC SIMILARITY).
CC -1- SUBUNIT: C8 IS COMPOSED OF THREE CHAINS: ALPHA, BETA AND GAMMA.
CC THE BETA CHAIN BINDS TO THE C8 ALPHA CHAIN AND TO THE C5B-C7
CC COMPLEX, PRESUMABLY TO C5B. IT IS ESSENTIAL TO THE INCORPORATION
CC OF C8 INTO THE C5B-C8 COMPLEX (BY SIMILARITY).
CC -1- SIMILARITY: TO COMPLEMENT FACTORS C6, C7, C9, AND TO PERFORIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF418597; AAL16647.1; -
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR002172; LDL_recept_A.
CC InterPro: IPR01862; MAC_perforin.
CC InterPro: IPR000884; TSP1.
CC Pfam: PF00057; ldl_recept_a; 1.
CC Pfam: PF01823; MACPF; 1.
CC SMART: SM00181; EGF; 1.
CC PROSITE: PS00022; EGF_1; 1.
CC PROSITE: PS01186; EGF_2; 1.
CC PROSITE: PS50068; LDLRA_2; 1.
CC PROSITE: PS00279; MAC_perforin; 1.
CC PROSITE: PS50092; TSP1; 2.
CC KW Complement pathway; Complement alternate pathway; Glycoprotein;
CC KW Signal; Plasma; Membrane attack complex; Cytolysis; EGF-like domain;
CC KW Repeat; Transmembrane.
CC FT SIGNAL 1 31 POTENTIAL.
CC FT PROPEP 32 44 BY SIMILARITY.
CC FT CHAIN 45 587 COMPLEMENT COMPONENT C8 BETA CHAIN.
CC FT DOMAIN 117 159 TSP TYPE-1.
CC FT DOMAIN 308 324 LDL-RECEPTOR CLASS A.
CC FT TRANSMEM 329 348 POTENTIAL.
CC FT DOMAIN 495 531 EGF-LIKE.
CC FT DOMAIN 541 587 TSP TYPE-1 2.
CC FT DISULFD 118 129 BY SIMILARITY.
CC FT DISULFD 123 142 BY SIMILARITY.
CC FT DISULFD 374 399 BY SIMILARITY.
CC SO SEQUENCE 587 AA; 65540 MW; 6D80417659FC879E CRC64;
OY 2 NSGDYKE---DYGEFAHTRADM 20
Db 325 NYGEYRQIYRDYGYTHTEATL 346

```

Search completed: June 16, 2003, 12:07:24
Job time : 3.33402 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:03:38 ; Search time 2.58197 Seconds
(Without alignments)
781.893 Million cell updates/sec

Title: US-09-986-682B-1
Perfect score: 114
Sequence: 1 MNSGDYKEDYGFAHTRADML 21

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	59.6	473	2	levanusucrase (EC 2
2	66	57.9	472	2	levanusucrase (EC 2
3	59	51.8	489	2	levanusucrase (Impo
4	55	48.2	646	2	hypothetical prote
5	48	42.1	310	2	UDP-n-acetylenolpy
6	48	42.1	378	2	hypothetical prote
7	48	42.1	478	2	6-phospho-beta-glu
8	47	41.2	289	2	hypothetical prote
9	47	41.2	567	2	lipoprotein (Impor
10	46	40.4	217	2	adenylate kinase (
11	46	40.4	365	2	hypothetical prote
12	46	40.4	677	2	hypothetical prote
13	45	39.5	608	2	hypothetical prote
14	45	39.5	728	2	probable very-long
15	45	39.5	729	2	hypothetical prote
16	45	39.5	749	2	probable mutator-1
17	45	39.5	882	2	hypothetical prote
18	45	39.5	895	2	mutator-like trans
19	45	39.5	942	2	mutator-like trans
20	45	39.5	942	2	mutator-like trans
21	45	39.5	1071	2	protein F27F5.15
22	45	39.5	1127	2	probable mutator-1
23	44	38.6	250	2	pentapeptide repa
24	44	38.6	364	2	hypothetical prote
25	44	38.6	484	2	leucine aminopepti
26	44	38.6	533	2	RNA-directed RNA p
27	43	37.7	227	2	cytidylate kinase
28	43	37.7	353	2	hypothetical prote
29	43	37.7	457	2	4-aminobutylate tr

30	43	37.7	478	2	6-phospho-beta-glu
31	43	37.7	958	1	glucan 1,4-alpha-g
32	43	37.7	1094	2	probable recb prot
33	43	37.7	1320	2	phosphoribosylform
34	43	37.7	1948	2	N conserved hypoch
35	42	36.8	143	2	hypothetical prote
36	42	36.8	193	1	hypothetical prote
37	42	36.8	193	1	adenylate kinase (
38	42	36.8	198	2	NAD(P)H oxidoreduc
39	42	36.8	204	2	uridine-monophosph
40	42	36.8	220	2	adenylate kinase (
41	42	36.8	321	2	probable succinate
42	42	36.8	332	2	pyruvate formate-1
43	42	36.8	406	2	probable glycosyl
44	42	36.8	477	2	beta-glucosidase (
45	42	36.8	477	2	beta-glucosidase (

ALIGNMENTS

RESULT 1
A25040
levanusucrase (EC 2.4.1.10) sacB precursor - Bacillus subtilis
M:Alternate names: sucrose 6-fructosyl-transferase
C:Species: Bacillus subtilis
C:Date: 16-Aug-1988 #sequence_revision 20-Feb-1995 #text_change 20-Jun-2000
C:Accession: S07309; A25040; I39967; H69702
R:Steinmetz, M.; le Coq, D.; Aymerich, S.; Gonzy-Treboul, G.; Gay, P.
Mol. Gen. Genet. 200, 220-228, 1985
A:Title: The DNA sequence of the gene for the secreted Bacillus subtilis enzyme levan
A:Reference number: S07309; MUID:81295507; PMID:2993818
A:Accession: S07309
A:Molecule type: DNA
A:Residues: 1-473 <STE>
A:Cross-references: EMBL:X02730; NID:940118; PIDN:CAA26513.1; PID:g732568
A:Experimental source: Marburg
R:Shimotsu, H.; Henner, D.J.
J. Bacteriol. 168, 380-388, 1986
A:Title: Modulation of Bacillus subtilis levanusucrase gene expression by sucrose and
A:Reference number: A25040; MUID:87008406; PMID:2428811
A:Accession: A25040
A:Molecule type: DNA
A:Residues: 1-68 <SHD>
A:Cross-references: GB:M14202; NID:g143485; PIDN:AAA22725.1; PID:g143486
R:Fouet, A.; Arnand, M.; Klier, A.; Rapoport, G.
Biochem. Biophys. Res. Commun. 119, 795-800, 1984
A:Title: Characterization of the precursor form of the exocellular levanusucrase from B
A:Reference number: I39967; MUID:84178454; PMID:6424671
A:Accession: I39967
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-11, 13-62 <RES>
A:Cross-references: GB:K01987; NID:g143483; PIDN:AAA22724.1; PID:g143484
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Buschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fuma, S.; Galizzi, A.; Gal
lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
Y. M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpiltra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipal, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zunsstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: H69702
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

```

OY      2  TSSGDYKEDYGFAMHTRADM 20
        ID:|||||:|||||:|
Db      30  NDMNKETIGVSHIRYIM 48

RESULT 4
109906
hypothetical protein T22A6.260 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C:Accession: T09906
R:Bayan, M.; Zimmermann, W.; Gruenissen, A.; Manbut, R.; Bancroft, I.; Mewes, H.W.;
submitted to the Protein Sequence Database, June 1999
A:Reference number: 216896
A:Accession: T09906
A:Molecule type: DNA
A:Residues: 1-646 <BEV>
A:Cross-references: EMBL:AL078637; GSPDB:GN00062; ATSP:T22A6.260
A:Experimental source: cultivar Columbia; BAC clone T22A6
C:Genetics:
A:Gene: ATSP:T22A6.260
A:Map position: 4
A:Introns: 15/3; 56/2; 76/2; 116/2; 156/2; 260/3; 313/3; 349/2; 379/3; 427/1; 479/2;

Query Match      48.2%  Score 55;  DB 2;  Length 646;
Best Local Similarity 69.2%  Pred. No. 0.94;
Matches 9;  Conservative 3;  Mismatches 1;  Indels 0;  Gaps 0;

OY      5  DYKEDYGFAMHTR 17
        ID:|:|:|:|
Db      506  DYKDDWFFAHWTR 518

RESULT 5
D1679
UDP-n-acetylenolpyruvoylglyucosamine reductase (murB) RP248 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: D1679
R:Anderson, S.G.E.; Zomerodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9823893
A:Accession: D1679
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-810 <AND>
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CA14710.1; PID:g3866
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: murB; RP248

Query Match      42.1%  Score 48;  DB 2;  Length 310;
Best Local Similarity 53.8%  Pred. No. 5.6;
Matches 7;  Conservative 3;  Mismatches 3;  Indels 0;  Gaps 0;

OY      4  GDYKEDYGFAMHIT 16
        ID:|:|:|:|
Db      24  GEYKKDYMLKHLIT 36

RESULT 6
D97741
hypothetical protein murB [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: D97741
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.;
Science 293, 2093-2098, 2001
A:Title: Mechanisms of evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: D97741

```


Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Serot, A.; Kuch, M.; Tamkoshi, A.; Tanaka, T.; Terpstra, P.; Tononi, A.; Tosato, V.; Uchiyama, T.; Walters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MVID:98044033; PMID:9384377
 A:Accession: E69583
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1217 <KUN>
 A:Cross-references: GB:299104; GB:AL009126; NID:92632267; PIDN:CA111913.1; PID:92632404
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: ack
 C:Function:
 A:Description: catalyzes reversible phosphorylation of AMP with ATP to form two ADP
 A:Note: magnesium required
 C:Superfamily: adenylate kinase
 C:Keywords: ATP; P-loop; phosphotransferase
 F:7-14/Region: nucleotide-binding motif A (P-loop) #status atypical
 F:80-85/Region: nucleotide-binding motif B #status atypical
 F:28,84/Active site: His, Asp #status predicted

Query Match 40.4%; Score 46; DB 2; Length 217;
 Best Local Similarity 61.5%; Pred. No. 7.9;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 8 EDYGFHITRADM 20
 ||||| ||: ||
 22 EDYGFHITRADM 34

RESULT 11
 A71048
 hypothetical protein PH1673 - *Pyrococcus horikoshii*
 C:Species: *Pyrococcus horikoshii*
 C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
 C:Accession: A71048
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hakiwa, Y.; Hino, Y.; Yamamoto, S.; Sekiguchi, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon.
 A:Reference number: A71000; MVID:98344137; PMID:9679194
 A:Accession: A71048
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1365 <KAM>
 A:Cross-references: GB:AP000006; NID:93236133; PIDN:BA30785.1; PID:93258102
 A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank C:Genetics:
 A:Gene: PH1673
 C:Superfamily: *Pyrococcus abyssi* hypothetical protein PAB0337

Query Match 40.4%; Score 46; DB 2; Length 365;
 Best Local Similarity 47.1%; Pred. No. 14;
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 2 NSGDYKEDYGFHITRA 18
 : ||| : ||| : ||
 DB 249 NSGDYKEDYGFHITRA 265

RESULT 12
 G84614
 hypothetical protein At2g22620 [imported] - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: G84614
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shua, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
 A:Reference number: A84420; MVID:20083487; PMID:10617197
 A:Accession: G84614
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-677 <STO>
 A:Cross-references: GB:AE002093; NID:94314359; PIDN:AA15570.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g22620
 A:Map position: 2

Query Match 40.4%; Score 46; DB 2; Length 677;
 Best Local Similarity 53.8%; Pred. No. 28;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 5 DYKEDYGFHITR 17
 ||: |||: ||
 DB 545 DYKEDYGFHITR 557

RESULT 13
 H83284
 probable very-long-chain acyl-CoA synthetase PA2893 [imported] - *Pseudomonas aeruginosa*
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: H83284
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L. Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.
 A:Reference number: A82950; MVID:20437337; PMID:10984043
 A:Accession: H83284
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-608 <STO>
 A:Cross-references: GB:AE004715; GB:AE004091; NID:99948977; PIDN:AA06281.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA2893
 C:Superfamily: *Mycobacterium tuberculosis* probable fadD6 protein; acetate-CoA ligase

Query Match 39.5%; Score 45; DB 2; Length 608;
 Best Local Similarity 45.0%; Pred. No. 36;
 Matches 9; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 2 NSGDYKEDYGFHITRADM 21
 ||: || ||| ||| ||
 DB 450 NSGDYKEDYGFHITRADM 469

RESULT 14
 D86233
 hypothetical protein [imported] - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: D86233
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.M.; Chung, M.K.; Comai, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Hulzar, L. Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Matzila, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
 A:Reference number: A86141; MVID:21016719; PMID:11130712
 A:Accession: D86233
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-728 <STO>

A:Cross-references: GB:AE005173; NID:92160180; PIDN:AAB60743.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 39.5%; Score 45; DB 2; Length 728;
 Best Local Similarity 53.8%; Pred. No. 44;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 5 DYKEDYGFHHTR 17
 ||:|:||||:|
 Db 597 DYRDWFFAHVPR 609

RESULT 15

G96559

hypothetical protein F5F19.7 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002

C:Accession: G96559

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venier, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G96559

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-729 <STO>

A:Cross-references: GB:AE005173; NID:94220448; PIDN:AD12675.1; GSPDB:GN00141

C:Genetics:

A:Gene: F5F19.7

A:Map position: 1

C:Superfamily: Arabidopsis thaliana hypothetical protein F26C24.7

Query Match 39.5%; Score 45; DB 2; Length 729;

Best Local Similarity 69.2%; Pred. No. 44;

Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 6 YKEDYGFHHTRA 18
 |||||
 Db 663 YLEDIGFEHWTRA 675

Search completed: June 16, 2003, 12:10:40
 Job time : 4.58197 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:07:33 : Search time 3.22746 Seconds
(Without alignments)
695.665 Million cell updates/sec

Title: US-09-986-682B-1
Perfect score: 114
Sequence: 1 MNSGDYKEDYGFHITRADML 21

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB pep:.*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB pep:.*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB pep:.*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB pep:.*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB pep:.*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB pep:.*
7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB pep:.*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB pep:.*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB pep:.*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB pep:.*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB pep:.*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB pep:.*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB pep:.*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB pep:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	100.0	21	US-09-986-682B-1	Sequence 1, Appli
2	114	100.0	455	US-09-986-682B-3	Sequence 3, Appli
3	48	42.1	310	US-09-991-211-8	Sequence 8, Appli
4	44	38.6	18	US-10-084-813-201	Sequence 201, App
5	44	38.6	18	US-10-084-813-424	Sequence 424, App
6	44	38.6	342	US-09-852-156-2	Sequence 2, Appli
7	44	38.6	342	US-09-940-063-2	Sequence 2, Appli
8	43	37.7	18	US-10-084-813-202	Sequence 202, App
9	43	37.7	18	US-10-084-813-425	Sequence 425, App
10	43	37.7	18	US-10-084-813-426	Sequence 426, App
11	43	37.7	18	US-10-084-813-427	Sequence 427, App
12	43	37.7	18	US-10-084-813-428	Sequence 428, App
13	42	36.8	143	US-09-928-457-39	Sequence 39, Appli
14	42	36.8	364	US-10-153-668-472	Sequence 472, App
15	42	36.8	366	US-10-153-668-472	Sequence 472, App
16	42	36.8	479	US-09-815-242-10288	Sequence 10288, A
17	42	36.8	566	US-09-797-500-2	Sequence 2, Appli
18	41	36.0	126	US-09-815-242-14024	Sequence 14024, A
19	41	36.0	242	US-09-764-868-1177	Sequence 1177, Ap

20	41	36.0	245	US-09-764-868-865	Sequence 865, App
21	41	36.0	386	US-09-764-868-866	Sequence 866, App
22	40.5	35.5	504	US-10-158-895-2	Sequence 2, Appli
23	40.5	35.5	504	US-10-123-427-2	Sequence 2, Appli
24	40.5	35.5	504	US-10-123-427-6	Sequence 43, Appli
25	40.5	35.5	513	US-10-158-895-43	Sequence 11, Appli
26	40.5	35.5	517	US-10-158-895-11	Sequence 6794, Ap
27	40	35.1	88	US-09-738-656-6794	Sequence 36935, A
28	39	34.2	60	US-09-864-761-36935	Sequence 48469, A
29	39	34.2	10	US-09-864-761-48469	Sequence 155, App
30	39	34.2	194	US-09-791-279-155	Sequence 13239, A
31	39	34.2	284	US-09-815-242-13239	Sequence 4, Appli
32	39	34.2	477	US-09-966-147-4	Sequence 9, Appli
33	39	34.2	684	US-09-823-240-9	Sequence 2, Appli
34	39	34.2	822	US-09-966-147-2	Sequence 5, Appli
35	39	34.2	847	US-09-924-859A-5	Sequence 5845, Ap
36	39	34.2	1016	US-09-815-242-5845	Sequence 18, Appli
37	38.5	33.8	904	US-10-121-988-18	Sequence 18, Appli
38	38.5	33.8	904	US-09-894-998-18	Sequence 19, Appli
39	38	33.3	25	US-09-839-666-19	Sequence 15, Appli
40	38	33.3	26	US-09-839-666-15	Sequence 1566, Ap
41	38	33.3	130	US-09-815-242-5156	Sequence 1057, Ap
42	38	33.3	193	US-09-925-300-1057	Sequence 1273, Ap
43	38	33.3	252	US-09-925-301-1273	Sequence 5719, Ap
44	38	33.3	478	US-09-815-242-5719	Sequence 12480, A
45	38	33.3	478	US-09-815-242-12480	

ALIGNMENTS

RESULT 1
US-09-986-682B-1
Sequence 1, Application US/09986682B
Patent No. US20020115182A1
GENERAL INFORMATION:
APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU
KENKYUJO
TITLE OF INVENTION: POLYPEPTIDE HAVING --FRUCTOPURANOSIDASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/986, 682B
FILING DATE: 06-NO. US20020115182A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170, 630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25, 618
REFERENCE/DOCKET NUMBER: TSUSAKI=2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
FRAGMENT TYPE: N-terminal fragment

SEQUENCE DESCRIPTION: SEQ ID NO: 1
US-09-986-682B-1

Query Match 100.0%; Score 114; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 1e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSGDYKEDYGFAMHTRADML 21
1 MNSGDYKEDYGFAMHTRADML 21

Db 1 MNSGDYKEDYGFAMHTRADML 21

RESULT 2
US-09-986-682B-3
Sequence 3, Application US/09986682B
Patent No. US20020115182A1
GENERAL INFORMATION:
APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU
KENKYUO.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/986,682B
FILING DATE: 06-NO. US20020115182A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI=2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-986-682B-3

Query Match 100.0%; Score 114; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 2.6e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSGDYKEDYGFAMHTRADML 21
1 MNSGDYKEDYGFAMHTRADML 21

Db 1 MNSGDYKEDYGFAMHTRADML 21

RESULT 3
US-09-991-211-8
Sequence 8, Application US/09991211
Patent No. US20020156585A1
GENERAL INFORMATION:
APPLICANT: Pharmacia & Upjohn
TITLE OF INVENTION: CRYSTALLIZATION AND STRUCTURE DETERMINATION OF

TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS
TITLE OF INVENTION: UDP-N-ACETYLENOLPYRUVYLGLUCOSAMINE REDUCTASE (S. aureus
FILE REFERENCE: 268 6241 0101
CURRENT APPLICATION NUMBER: US/09/991,211
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 09/632,947
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/147,164
PRIOR FILING DATE: 1999-08-04
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 8
LENGTH: 310
TYPE: PRT
ORGANISM: Rickettsia prowazekii
US-09-991-211-8

Query Match 42.1%; Score 48; DB 9; Length 310;
Best Local Similarity 53.8%; Pred. No. 6.4;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 GDYKEDYGFAMHIT 16
1 MNSGDYKEDYGFAMHIT 16

Db 24 GEYKDYMLKHLT 36

RESULT 4
US-10-084-813-201
Sequence 201, Application US/10084813
Publication No. US20030068615A1
GENERAL INFORMATION:
APPLICANT: SAXINGER, CARL
TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE
FILE REFERENCE: 215875
CURRENT APPLICATION NUMBER: US/10/084,813
CURRENT FILING DATE: 2002-02-27
PRIOR APPLICATION NUMBER: PCT/US00/23505
PRIOR FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: US 60/151,270
PRIOR FILING DATE: 1999-08-27
NUMBER OF SEQ ID NOS: 1242
SOFTWARE: Patentln version 3.1
SEQ ID NO 201
LENGTH: 18
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: binding peptide
US-10-084-813-201

Query Match 38.6%; Score 44; DB 9; Length 18;
Best Local Similarity 61.5%; Pred. No. 1.3;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNSGDYKEDYGF 13
1 MAEDHYHEDYGS 13

Db 1 MAEDHYHEDYGS 13

RESULT 5
US-10-084-813-424
Sequence 424, Application US/10084813
Publication No. US20030068615A1
GENERAL INFORMATION:
APPLICANT: SAXINGER, CARL
TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE
FILE REFERENCE: 215875
CURRENT APPLICATION NUMBER: US/10/084,813
CURRENT FILING DATE: 2002-02-27
PRIOR APPLICATION NUMBER: PCT/US00/23505

PRIOR FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: US 60/151,270
PRIOR FILING DATE: 1999-08-27
NUMBER OF SEQ ID NOS: 1242
SOFTWARE: PatentIn version 3.1
SEQ ID NO 424
LENGTH: 18
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: binding peptide
US-10-084-813-424

Query Match 38.6%; Score 44; DB 9; Length 18;
Best Local Similarity 61.5%; Pred. No. 1.3;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFA 13
Db 1 MAEHDYHEDYGFS 13

RESULT 6
US-09-852-156-2
Sequence 2, Application US/09852156
Patent No. US20020076694A1
GENERAL INFORMATION:
APPLICANT: Littmap, Dan R.
Deng, Hongkui
Unutlma, Derya
Ramani, Vineet N.K.

TITLE OF INVENTION: NOVEL ALTERNATIVE G-COUPLED RECEPTORS
ASSOCIATED WITH RETROVIRAL ENTRY INTO CELLS, METHODS OF
IDENTIFYING THE SAME, AND DIAGNOSTIC AND THERAPEUTIC USES
THEREOF

NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/852,156
FILING DATE: 09-May-2001
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1049-1-009NCIP

TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-852-156-2

Query Match 38.6%; Score 44; DB 10; Length 342;
Best Local Similarity 61.5%; Pred. No. 31;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFA 13
Db 1 MAEHDYHEDYGFS 13

RESULT 7
US-09-940-063-2
Sequence 2, Application US/09940063
Patent No. US20020090657A1
GENERAL INFORMATION:
APPLICANT: Briskin, Michael J.
APPLICANT: Murphy, Kristine E.
APPLICANT: Wilbanks, Alyson M.
APPLICANT: Wu, Lijun
TITLE OF INVENTION: No. US20020090657A1el Antibodies and Ligands for "Bonzo"
FILE REFERENCE: 1855,1070-000
CURRENT APPLICATION NUMBER: US/09/940,063
CURRENT FILING DATE: 2001-08-27
PRIOR APPLICATION NUMBER: 09/449,437
PRIOR FILING DATE: 1999-11-24
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 342
TYPE: PRT
ORGANISM: Homo sapiens
US-09-940-063-2

Query Match 38.6%; Score 44; DB 10; Length 342;
Best Local Similarity 61.5%; Pred. No. 31;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFA 13
Db 1 MAEHDYHEDYGFS 13

RESULT 8
US-10-084-813-202
Sequence 202, Application US/10084813
Publication No. US20030068615A1
GENERAL INFORMATION:
APPLICANT: SAXINGER, CARL
TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE
FILE REFERENCE: 215875

CURRENT APPLICATION NUMBER: US/10/084,813
CURRENT FILING DATE: 2002-02-27

PRIOR APPLICATION NUMBER: PCT/US00/23505
PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: US-60/151,270
PRIOR FILING DATE: 1999-08-27

NUMBER OF SEQ ID NOS: 1242
SOFTWARE: PatentIn version 3.1
SEQ ID NO 202

LENGTH: 18
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: binding peptide
US-10-084-813-202

Query Match 37.7%; Score 43; DB 9; Length 18;
Best Local Similarity 77.8%; Pred. No. 1.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 5 DYKEDYGFA 13
Db 1 MAEHDYHEDYGFS 13

Db 1 DYHEDYGFS 9

RESULT 9

US-10-084-813-425

Sequence 425, Application US/10084813
Publication No. US20030068615A1

GENERAL INFORMATION:

APPLICANT: SAXINGER, CARL

TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE

FILE REFERENCE: 215875

CURRENT APPLICATION NUMBER: US/10/084,813

CURRENT FILING DATE: 2002-02-27

PRIOR APPLICATION NUMBER: PCT/US00/23505

PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: US 60/151,270

PRIOR FILING DATE: 1999-08-27

NUMBER OF SEQ ID NOS: 1242

SOFTWARE: PatentIn version 3.1

SEQ ID NO 425

LENGTH: 18

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: binding peptide

US-10-084-813-425

Query Match 37.7%; Score 43; DB 9; Length 18;

Best Local Similarity 77.8%; Pred. No. 1.9;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 DYKEDYGFA 13

Db 4 DYHEDYGFS 12

RESULT 10

US-10-084-813-426

Sequence 426, Application US/10084813

Publication No. US20030068615A1

GENERAL INFORMATION:

APPLICANT: SAXINGER, CARL

TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE

FILE REFERENCE: 215875

CURRENT APPLICATION NUMBER: US/10/084,813

CURRENT FILING DATE: 2002-02-27

PRIOR APPLICATION NUMBER: PCT/US00/23505

PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: US 60/151,270

PRIOR FILING DATE: 1999-08-27

NUMBER OF SEQ ID NOS: 1242

SOFTWARE: PatentIn version 3.1

SEQ ID NO 426

LENGTH: 18

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: binding peptide

US-10-084-813-426

Query Match 37.7%; Score 43; DB 9; Length 18;

Best Local Similarity 77.8%; Pred. No. 1.9;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 DYKEDYGFA 13

Db 3 DYHEDYGFS 11

RESULT 11

US-10-084-813-427

Sequence 427, Application US/10084813
Publication No. US20030068615A1

GENERAL INFORMATION:

APPLICANT: SAXINGER, CARL

TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE

FILE REFERENCE: 215875

CURRENT APPLICATION NUMBER: US/10/084,813

CURRENT FILING DATE: 2002-02-27

PRIOR APPLICATION NUMBER: PCT/US00/23505

PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: US 60/151,270

PRIOR FILING DATE: 1999-08-27

NUMBER OF SEQ ID NOS: 1242

SOFTWARE: PatentIn version 3.1

SEQ ID NO 427

LENGTH: 18

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: binding peptide

US-10-084-813-427

Query Match 37.7%; Score 43; DB 9; Length 18;

Best Local Similarity 77.8%; Pred. No. 1.9;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 DYKEDYGFA 13

Db 2 DYHEDYGFS 10

RESULT 12

US-10-084-813-428

Sequence 428, Application US/10084813

Publication No. US20030068615A1

GENERAL INFORMATION:

APPLICANT: SAXINGER, CARL

TITLE OF INVENTION: POLYPEPTIDES THAT BIND HIV GP120 AND RELATED NUCLEIC
ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE

FILE REFERENCE: 215875

CURRENT APPLICATION NUMBER: US/10/084,813

CURRENT FILING DATE: 2002-02-27

PRIOR APPLICATION NUMBER: PCT/US00/23505

PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: US 60/151,270

PRIOR FILING DATE: 1999-08-27

NUMBER OF SEQ ID NOS: 1242

SOFTWARE: PatentIn version 3.1

SEQ ID NO 428

LENGTH: 18

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: binding peptide

US-10-084-813-428

Query Match 37.7%; Score 43; DB 9; Length 18;

Best Local Similarity 77.8%; Pred. No. 1.9;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 DYKEDYGFA 13

Db 1 DYHEDYGFS 9

RESULT 13

US-09-928-457-39

Sequence 39, Application US/09928457

Patent No. US20020164603A1

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: DNA, specific proteins and peptides

```

TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method
TITLE OF INVENTION: for obtaining them and their biological application.
NUMBER OF SEQUENCES: 99
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30 (OEB)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/928,457
FILING DATE: 2001-08-14
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/214,759
FILING DATE: 199-12-10
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 143 acids amin,s
TYPE: acide amin,
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..143
US-09-928-457-39

```

```

Query Match          36.8%; Score 42; DB 9; Length 143;
Best Local Similarity 53.3%; Pred. No. 26;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

OY      4 GDYKEDYGFAHITRA 18
        |||:| ||:| |
Db      67 GDYEEVNGFEYIDKA 81

```

```

RESULT 14
US-10-153-668-472
; Sequence 472, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 472
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-472

```

```

Query Match          36.8%; Score 42; DB 9; Length 364;
Best Local Similarity 46.7%; Pred. No. 69;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

OY 5 DYKEDYGFAHITRAD 19

```

Db      108 DIVKDYAFVHMERAE 122

```

```

RESULT 15
US-10-153-668-228
; Sequence 228, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 228
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-228

```

```

Query Match          36.8%; Score 42; DB 9; Length 366;
Best Local Similarity 46.7%; Pred. No. 69;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

OY      5 DYKEDYGFAHITRAD 19
        |||:| ||:| |
Db      108 DIVKDYAFVHMERAE 122

```

```

Search completed: June 16, 2003, 12:22:35
Job time : 5.22746 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:04:38 ; Search time 1.89344 Seconds
(without alignments)
326.327 Million cell updates/sec

Title: US-09-986-682B-1

Perfect score: 114

Sequence: 1 MNSGDYKEDYGFATRADML 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	100.0	21	2	US-08-870-827-1
2	114	100.0	21	2	US-09-317-179-1
3	114	100.0	435	2	US-08-870-827-3
4	114	100.0	455	4	US-09-317-179-3
5	48	42.1	310	4	US-09-633-947B-8
6	45.5	39.9	133	4	US-09-134-001C-4596
7	44	38.6	342	2	US-08-742-011-2
8	44	38.6	342	4	US-09-275-384B-5
9	44	38.6	342	4	US-09-116-498-2
10	44	38.6	342	4	US-09-449-437A-2
11	44	38.6	342	4	US-09-517-605-9
12	43	37.7	425	4	US-09-134-001C-4433
13	43	37.7	425	4	US-08-073-383-4
14	42	36.8	566	1	US-08-428-415-4
15	42	36.8	566	1	US-08-379-685-4
16	42	36.8	566	3	US-08-854-029-4
17	42	36.8	566	3	US-08-848-810-2
18	42	36.8	566	4	US-08-428-762-4
19	42	36.8	566	5	PCT-US94-06365-4
20	41	36.0	197	2	US-08-879-561-5
21	40.5	35.5	504	2	US-08-752-891-2
22	40.5	35.5	504	2	US-08-752-891-6
23	40.5	35.5	504	2	US-09-144-178-2
24	40.5	35.5	504	2	US-09-144-178-6
25	40.5	35.5	504	4	US-09-406-854-2
26	40.5	35.5	504	4	US-09-406-854-6
27	40.5	35.5	504	4	US-09-529-279-2

28	40.5	35.5	513	4	US-09-529-279-43	Sequence 43, Appli
29	40.5	35.5	517	4	US-09-529-279-11	Sequence 11, Appli
30	39.5	34.6	891	1	US-08-042-747A-6	Sequence 6, Appli
31	39	34.2	194	2	US-08-879-561-11	Sequence 11, Appli
32	39	34.2	477	2	US-08-359-705B-4	Sequence 4, Appli
33	39	34.2	477	2	US-08-286-846A-4	Sequence 4, Appli
34	39	34.2	477	2	US-08-457-880A-4	Sequence 4, Appli
35	39	34.2	477	2	US-08-444-622A-4	Sequence 4, Appli
36	39	34.2	477	3	US-08-942-562-4	Sequence 4, Appli
37	39	34.2	477	4	US-09-156-923-4	Sequence 4, Appli
38	39	34.2	821	1	US-08-339-578-2	Sequence 2, Appli
39	39	34.2	822	2	US-08-359-705B-2	Sequence 2, Appli
40	39	34.2	822	2	US-08-286-846A-2	Sequence 2, Appli
41	39	34.2	822	2	US-08-457-880A-2	Sequence 2, Appli
42	39	34.2	822	3	US-08-444-622A-2	Sequence 2, Appli
43	39	34.2	822	3	US-08-942-562-2	Sequence 2, Appli
44	39	34.2	822	3	US-09-156-923-2	Sequence 2, Appli
45	39	34.2	847	1	US-08-286-305A-5	Sequence 5, Appli

ALIGNMENTS

```
RESULT 1
US-08-870-827-1
; Sequence 1, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: Tsusaki et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; NUMBER OF INVENTIONS: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: BROWDY AND NEWMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI=2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; FRAGMENT TYPE: N-terminal fragment
; US-08-870-827-1

Query Match 100.0%; Score 114; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.4e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 MNSGDYKEDYGFATRADML 21
DB 1 MNSGDYKEDYGFATRADML 21

RESULT 2
US-09-317-179-1
Sequence 1, Application US/09317179
Patent No. 6383769
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/09/317,179
FILING DATE: 24-May-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/870,827
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
FRAGMENT TYPE: N-terminal fragment
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-317-179-1

Query Match 100.0%; Score 114; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.4e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSGDYKEDYGFAHITRADML 21
DB 1 MNSGDYKEDYGFAHITRADML 21

RESULT 3
US-08-870-827-3
Sequence 3, Application US/08870827
Patent No. 5962297
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,827
FILING DATE: 06-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-870-827-3

Query Match 100.0%; Score 114; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSGDYKEDYGFAHITRADML 21
DB 1 MNSGDYKEDYGFAHITRADML 21

RESULT 4
US-09-317-179-3
Sequence 3, Application US/09317179
Patent No. 6383769
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/317,179
FILING DATE: 24-May-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/870,827
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-317-179-3

Query Match 100.0%; Score 114; DB 4; Length 455;
Best Local Similarity 100.0%; Pred. No. 1,1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MNSGDYKEDYGFATIRADM 21
Db 1 MNSGDYKEDYGFATIRADM 21

RESULT 5
US-09-632-947B-8

Sequence 8, Application US/09632947B
Patent No. 6356845

GENERAL INFORMATION:

APPLICANT: Pharmacia & Upjohn

TITLE OF INVENTION: CRYSTALLIZATION AND STRUCTURE DETERMINATION OF

TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS

TITLE OF INVENTION: UDP-N-ACETYLENOLPYRUVYLGLUCOSAMINE REDUCTASE (S. aureus)

FILE REFERENCE: 268,6241 0101

CURRENT APPLICATION NUMBER: US/09/632,947B

CURRENT FILING DATE: 2000-08-04

PRIOR APPLICATION NUMBER: 60/147,164

PRIOR FILING DATE: 1999-08-04

NUMBER OF SEQ ID NOS: 11

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 8

LENGTH: 310

TYPE: PRT

ORGANISM: Rickettsia prowazekii

US-09-632-947B-8

Query Match 42.1%; Score 48; DB 4; Length 310;
Best Local Similarity 53.8%; Pred. No. 3.5;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 4 GDYKEDYGFATIT 16

Db 24 GEYKKDYNLKHLT 36

RESULT 6
US-09-134-001C-4596

Sequence 4596, Application US/09134001C
Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4596

LENGTH: 133

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4596

Query Match 39.9%; Score 45.5; DB 4; Length 133;
Best Local Similarity 34.5%; Pred. No. 3.4;
Matches 10; Conservative 5; Mismatches 5; Indels 9; Gaps 1;

Oy 1 MNSGD-----YKEDYGFATIRADM 20

Db 54 LNDGSETPNVEIKYDKKSKSHLEKADL 82

RESULT 7
US-08-742-011-2

Sequence 2, Application US/08742011
Patent No. 5824504

GENERAL INFORMATION:

APPLICANT: Eishourdayy, Nabil A.

APPLICANT: Bergsma, Derek J.

TITLE OF INVENTION: Human 7-Transmembrane Receptor

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19406-2799

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/742,011

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Han, William T.

REGISTRATION NUMBER: 34,344

REFERENCE/DOCKET NUMBER: ATG50020P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5219

TELEFAX: 610-270-5090

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 342 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-742-011-2

Query Match 38.6%; Score 44; DB 2; Length 342;
Best Local Similarity 61.5%; Pred. No. 18;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 1 MNSGDYKEDYGF 13

Db 1 MAEHYHEDYGF 13

RESULT 8
US-09-275-384B-5

Sequence 5, Application US/09275384B
Patent No. 6232084

GENERAL INFORMATION:

APPLICANT: MACPHER, COLIN HOUSTON

APPLICANT: MOORES, KITTY

TITLE OF INVENTION: NEW USE

FILE REFERENCE: GH-31106

CURRENT APPLICATION NUMBER: US/09/275,384B

CURRENT FILING DATE: 1999-03-24

PRIOR APPLICATION NUMBER: 9806677.2

PRIOR FILING DATE: 1998-03-27

NUMBER OF SEQ ID NOS: 9

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 5

LENGTH: 342

TYPE: PRT

ORGANISM: HOMO SAPIENS

US-09-275-384B-5

Query Match 38.6%; Score 44; DB 4; Length 342;
Best Local Similarity 61.5%; Pred. No. 18;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFA 13
1 111111:
Db 1 MAEHYHEDYGF 13

RESULT 9
US-09-116-498-2
Sequence 2, Application US/09116498
Patent No. 6251582

GENERAL INFORMATION:

APPLICANT: Littman, Dan R.

Deng, Hongkui

Unulmaz, Derya

Ramani, Vineet N.R.

TITLE OF INVENTION: NOVEL ALTERNATIVE G-COUPLED RECEPTORS

ASSOCIATED WITH RETROVIRAL ENTRY INTO CELLS, METHODS OF

IDENTIFYING THE SAME, AND DIAGNOSTIC AND THERAPEUTIC USES

THEREOF

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: David A. Jackson, Esq.

STREET: 411 Hackensack Ave, Continental Plaza, 4th

Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/116,498

FILING DATE: 16-Jul-1998

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 1049-1-009

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 342 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHEtical: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-116-498-2

Query Match 38.6%; Score 44; DB 4; Length 342;
Best Local Similarity 61.5%; Pred. No. 18;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFA 13
1 111111:
Db 1 MAEHYHEDYGF 13

RESULT 10
US-09-449-437A-2

Sequence 2, Application US/09449437A

Patent No. 6319675

GENERAL INFORMATION:

APPLICANT: Briskin, Michael J.

APPLICANT: Murphy, Kristine E.

APPLICANT: Wilbanks, Alyson M.

APPLICANT: Wu, Lijun

TITLE OF INVENTION: No. 6319675el Antibodies and Ligands for "Bonzo"

TITLE OF INVENTION: Chemokine Receptor

FILE REFERENCE: 1855.1070-000

CURRENT APPLICATION NUMBER: US/09/449,437A

CURRENT FILING DATE: 2001-01-09

NUMBER OF SEQ ID NOS: 18

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 342

TYPE: PRT

ORGANISM: Homo sapiens

US-09-449-437A-2

Query Match 38.6%; Score 44; DB 4; Length 342;
Best Local Similarity 61.5%; Pred. No. 18;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFA 13
1 111111:
Db 1 MAEHYHEDYGF 13

RESULT 11
US-09-517-605-9
Sequence 9, Application US/09517605
Patent No. 6391567

GENERAL INFORMATION:

APPLICANT: Littman, Dan R.

APPLICANT: Kwon, Douglas S.

APPLICANT: van Kooyk, Yvette

APPLICANT: Geltenbeck, Theo

TITLE OF INVENTION: METHODS OF USING A FACILITATOR OF RETROVIRAL ENTRY INTO

TITLE OF INVENTION: CELLS

FILE REFERENCE: 1049-1-017

CURRENT APPLICATION NUMBER: US/09/517,605

CURRENT FILING DATE: 2000-03-02

NUMBER OF SEQ ID NOS: 17

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 9

LENGTH: 342

TYPE: PRT

ORGANISM: Homo sapiens

US-09-517-605-9

Query Match 38.6%; Score 44; DB 4; Length 342;
Best Local Similarity 61.5%; Pred. No. 18;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFA 13
1 111111:
Db 1 MAEHYHEDYGF 13

RESULT 12
US-09-134-001C-4433
Sequence 4433, Application US/09134001C
Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucelte-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO: 4433
LENGTH: 425
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4433

Query Match 37.7%; Score 43; DB 4; Length 425;
Best Local Similarity 61.5%; Pred. No. 33;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 9 DYGFHITRADML 21
| | | | | | | | | |
DB 5 DKGFHILTRKDL 17

RESULT 13
US-08-073-383-4
Sequence 4, Application US/08073383
Patent No. 5443962
GENERAL INFORMATION:
APPLICANT: Draetta, Giulio
APPLICANT: Cottarel, Guillaume
APPLICANT: Damaguez, Veronique
TITLE OF INVENTION: {fillin "Insert Title of Application" (ASSAY AND REAGENTS FOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/073,383
FILING DATE: 19930604
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MII-009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
TITLE OF INVENTION: Cold Spring Harbor Laboratory
SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-073-383-4

Query Match 36.8%; Score 42; DB 1; Length 566;
Best Local Similarity 47.1%; Pred. No. 66;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 GDYKEDYGFHITRADM 20
| | | | | | | | | |
DB 249 GDTEDDGFVILSDL 265

RESULT 14

US-08-428-415-4
Sequence 4, Application US/08428415
Patent No. 5756335

GENERAL INFORMATION:
APPLICANT: Cold Spring Harbor Laboratory
TITLE OF INVENTION: No. 5756335el Human cdc25 Genes, Encoded
TITLE OF INVENTION: Products and Uses Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,415
FILING DATE: 24 April 1995

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MII-019CP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-428-415-4

Query Match 36.8%; Score 42; DB 1; Length 566;
Best Local Similarity 47.1%; Pred. No. 66;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 GDYKEDYGFHITRADM 20
| | | | | | | | | |
DB 249 GDTEDDGFVILSDL 265

RESULT 15
US-08-379-685-4
Sequence 4, Application US/08379685
Patent No. 5770423
GENERAL INFORMATION:
APPLICANT: Cold Spring Harbor Laboratory
TITLE OF INVENTION: No. 5770423el Human cdc25 Genes, Encoded
TITLE OF INVENTION: Products and Uses Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/379,685
FILING DATE: 26 January 1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MI-019-DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-379-685-4

Query Match 36.8%; Score 42; DB 1; Length 566;
Best Local Similarity 47.1%; Pred. No. 66;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
Oy 4 GDYKEDYGFAMHTRADM 20
||:|||||:|:
Db 249 GDTEDDGFVDLLESDL 265

Search completed: June 16, 2003, 12:11:31
Job time : 2.89344 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:55:48 ; Search time 5.59426 seconds
(without alignments)
500.202 Million cell updates/sec

Title: US-09-986-682B-1

Perfect score: 114

Sequence: 1 MNSGDYKEDYGFHITRADML 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	100.0	21	19	AAW48307
2	114	100.0	487	19	AAW48306
3	68	55.6	62	11	AAW48307
4	66	57.9	137	22	AAW48307
5	66	57.9	165	22	ABG04300
6	66	57.9	175	22	ABG04304
7	66	57.9	194	22	ABG26186
8	66	57.9	231	22	ABG04299
9	66	57.9	241	22	ABG21816
10	66	57.9	242	22	ABG04290

11	66	57.9	256	22	ABG04297	Novel human diago
12	66	57.9	266	22	AAU30761	Novel human secret
13	66	57.9	267	22	ABG26189	Novel human diago
14	66	57.9	268	22	ABG10291	Novel human diago
15	66	57.9	276	22	ABG04324	Novel human diago
16	66	57.9	280	22	ABG14490	Novel human diago
17	66	57.9	282	22	ABG24525	Novel human diago
18	66	57.9	287	22	ABG21864	Novel human diago
19	66	57.9	288	22	ABG04293	Novel human diago
20	66	57.9	293	22	ABG05672	Novel human diago
21	66	57.9	306	22	ABG04301	Novel human diago
22	66	57.9	314	22	ABG04303	Novel human diago
23	66	57.9	314	22	ABG10673	Novel human diago
24	66	57.9	315	22	ABG26187	Novel human diago
25	66	57.9	324	22	ABG10297	Novel human diago
26	66	57.9	331	22	ABG04296	Novel human diago
27	66	57.9	338	22	ABG21839	Novel human diago
28	66	57.9	342	22	ABG04302	Novel human diago
29	66	57.9	342	22	ABG10300	Novel human diago
30	66	57.9	343	22	ABG04315	Novel human diago
31	66	57.9	343	22	ABG26188	Novel human diago
32	66	57.9	347	22	ABG18055	Novel human diago
33	66	57.9	347	22	ABG26458	Novel human diago
34	66	57.9	350	22	ABG04291	Novel human diago
35	66	57.9	350	22	ABG13458	Novel human diago
36	66	57.9	350	22	ABG21691	Novel human diago
37	66	57.9	357	22	ABG10299	Novel human diago
38	66	57.9	360	22	ABG10298	Novel human diago
39	66	57.9	361	22	ABG26191	Novel human diago
40	66	57.9	369	22	ABG21840	Novel human diago
41	66	57.9	374	22	ABG04285	Novel human diago
42	66	57.9	376	22	ABG10307	Novel human diago
43	66	57.9	378	22	ABG07324	Novel human diago
44	66	57.9	378	22	ABG21624	Novel human diago
45	66	57.9	384	22	ABG10303	Novel human diago

ALIGNMENTS

RESULT 1
ID AAW48307 standard; Protein; 21 AA.
XX
AC AAW48307;
XX
DT 04-JUN-1998 (first entry)
XX
DE Bacillus sp. V230 beta-fructofuranosidase N-terminal fragment.
KW Beta-fructofuranosidase; xylosyl fructoside; isomaltosyl fructoside;
KW lactosucrose; fructosyltrehalose; anticariogenic; growth promoter;
KW Bifidobacteria; mineral absorption promoting activity; taste; texture;
KW cosmetic; pharmaceutical; sweetener.
XX
OS Bacillus sp.
XX
PN EP812915-A2.
XX
PD 17-DEC-1997.
XX
PF 10-JUN-1997; 97EP-0304032.
XX
PR 10-JUN-1996; 96JP-0170630.
XX
PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
PI Chaen H. Kubota M. Tsusaki K.
DR WPI; 1998-034976/04.
XX
PT Bacillus beta-fructofuranosidase enzyme - useful for producing
fructo:furanosylated saccharide(s) or alcohol(s)

XX Claim 1; Page 19; 32pp; English.
PS
XX This is an N-terminal fragment of beta-fructofuranosidase from
CC Bacillus sp. V230, an enzyme that has an optimum temperature of 50 deg. C
CC at pH 6. This enzyme catalyses fructofuranosyl transfer reactions from a
CC fructofuranosyl donor to a fructofuranosyl acceptor where the donor is
CC sucrose, raffinose or erlose and the acceptor is selected from alcohols,
CC sugar alcohols and saccharides having no beta-fructofuranosidic linkages,
CC especially where the reaction product is xylosyl fructoside, erlose,
CC isomaltosyl fructoside, lactosucrose or fructosyltrehalose.
CC Such reaction products have a satisfactory taste and sweetness, a
CC moderate viscosity and humectancy, an effective anticariogenic activity,
CC growth promoting activity for Bifidobacteria, mineral-absorption
CC promoting activity. The products can be used to improve the tastes and
CC textures of food products, cosmetics and pharmaceuticals and are useful
CC as sweeteners in the food and pharmaceutical industries.

XX Sequence 21 AA;

Query Match 100.0%; Score 114; DB 19; Length 21;
Best Local Similarity 100.0%; Pred. No. 3e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFPAHITRADML 21
DB 1 MNSGDYKEDYGFPAHITRADML 21

RESULT 2

AAW48306
ID AAW48306 standard; Protein; 487 AA.

XX AAW48306;

DT 04-JUN-1998 (first entry)

XX Bacillus sp. V230 beta-fructofuranosidase.

XX Beta-fructofuranosidase; xylosyl fructoside; isomaltosyl fructoside;
XX lactosucrose; fructosyltrehalose; anticariogenic; growth promoter;

XX Bifidobacteria; mineral-absorption promoting activity; taste; texture;
XX cosmetic; pharmaceutical; sweetener.

XX Bacillus sp.

XX Key Location/Qualifiers

XX Peptide 1..32

XX Protein 33..487

XX /label= beta-fructofuranosidase

XX EP812915-A2.

XX 17-DEC-1997.

XX 10-JUN-1997; 97EP-0304032.

XX 10-JUN-1996; 96JP-0170630.

XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

XX Chaen H, Kubota M, Tsusaki K;

XX WPI: 1998-034976/04.

XX N-PSDB: AAV17621.

XX Bacillus beta-fructofuranosidase enzyme - useful for producing

XX fructo:furanosylated saccharide(s) or alcohol(s)

XX Claim 1; Pages 23-25; 32pp; English.

CC This is beta-fructofuranosidase from Bacillus sp. V230, an enzyme that
CC has an optimum temperature of 50 deg. C at pH 6. It catalyses
CC fructofuranosyl transfer reactions from a fructofuranosyl donor to a
CC fructofuranosyl acceptor where the donor is sucrose, raffinose or erlose
CC and the acceptor is selected from alcohols, sugar alcohols and
CC saccharides having no beta-fructofuranosidic linkages,
CC especially where the reaction product is xylosyl fructoside, erlose,
CC isomaltosyl fructoside, lactosucrose or fructosyltrehalose.
CC Such reaction products have a satisfactory taste and sweetness, a
CC moderate viscosity and humectancy, an effective anticariogenic activity,
CC growth promoting activity for Bifidobacteria, mineral-absorption
CC promoting activity. The products can be used to improve the tastes and
CC textures of food products, cosmetics and pharmaceuticals and are useful
CC as sweeteners in the food and pharmaceutical industries.

XX Sequence 487 AA;

Query Match 100.0%; Score 114; DB 19; Length 487;
Best Local Similarity 100.0%; Pred. No. 9.2e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNSGDYKEDYGFPAHITRADML 21
DB 33 MNSGDYKEDYGFPAHITRADML 53

RESULT 3

AAW05992
ID AAW05992 standard; Protein; 62 AA.

XX AAW05992;

DT 21-NOV-1990 (first entry)

XX Signal peptide and N-terminal of mature levansucrase.

XX Levansucrase; sacb gene; signal peptide; interferon; dss.

XX Bacillus subtilis.

XX Key Location/Qualifiers

XX Peptide 1..29

XX Cleavage-site 29..30

XX Protein 30..62

XX /label=Levansucrase protein

XX /note="N-terminal"

XX CA1270779-A.

XX 26-JUN-1990.

XX 09-NOV-1984; 84CA-0467509.

XX 09-NOV-1984; 84CA-0467509.

XX (INSP) INST PASTEUR.

XX FOUDET A, KLIER A, RAPOPORT G, DEDONDER R, ARNAUD M;

XX WPI: 1990-217285/29.

XX N-PSDB: AA005295.

XX New DNA encoding the signal peptide or Bacillus subtilis sacb

XX gene - inducing secretion of heterologous proteins expressed in

XX transformed cells.

XX Disclosure: ; pp; English.

XX Signal sequence may be fused to sequence encoding mature peptide

XX in a B.subtilis expression system, allowing the desired peptide

XX (eg. interferon, levansucrase) to be collected in cell secretions.

XX

Seq	Sequence	62 AA:	59-68:	Score 68:	DB 11:	Length 62:
QY	Query Match			59.68:		
	Best Local Similarity			65.08:		
	Matches 13: Conservative			1: Mismatches	6: Indels	0: Gaps
						0:
Db	2 NSGDKEDYGFATHTRADML 21					
	1 111 11 1111 111					
	33 NQPKETTYGISHTRDML 52					
RESULT 4						
ID	ABG04300					
AC	ABG04300 standard; Protein: 137 AA.					
XX	ABG04300:					
XX	13-FEB-2002 (first entry)					
XX	Novel human diagnostic protein #4291.					
XX	Human: chromosome mapping; gene mapping; gene therapy; forensic;					
XX	food supplement; medical imaging; diagnostic; genetic disorder.					
XX	Homo sapiens.					
XX	MO200175067-A2.					
XX	11-OCT-2001.					
XX	30-MAR-2001; 2001WO-US08631.					
XX	31-MAR-2000; 2000US-0540217.					
XX	23-AUG-2000; 2000US-0649167.					
XX	(HYSE-) HYSEQ INC.					
XX	Dermanac RT, Liu C, Tang YT;					
XX	WPI; 2001-639362/73.					
XX	N-PSDB; AAS68487.					
XX	New isolated polynucleotide and encoded polypeptides, useful in					
XX	diagnostics, forensics, gene mapping, identification of mutations					
XX	responsible for genetic disorders or other traits and to assess					
XX	biodiversity					
XX	Claim 20; SEQ ID No 34659; 103pp; English.					
XX	The invention relates to isolated polynucleotide (I) and					
XX	polypeptide (II) sequences. (I) is useful as hybridisation probes,					
XX	polymerase chain reaction (PCR) primers, oligomers, and for chromosome					
XX	and gene mapping, and in recombinant production of (II). The					
XX	polynucleotides are also used in diagnostics as expressed sequence tags					
XX	for identifying expressed genes. (I) is useful in gene therapy techniques					
XX	to restore normal activity of (II) or to treat disease states involving					
XX	(II). (II) is useful for generating antibodies against it, detecting or					
XX	quantitating a polypeptide in tissue, as molecular weight markers and as					
XX	a food supplement. (II) and its binding partners are useful in medical					
XX	imaging of sites expressing (II). (I) and (II) are useful for treating					
XX	disorders involving aberrant protein expression or biological activity.					
XX	The polypeptide and polynucleotide sequences have applications in					
XX	diagnostics, forensics, gene mapping, identification of mutations					
XX	responsible for genetic disorders or other traits to assess biodiversity					
XX	and to produce other types of data and products dependent on DNA and					
XX	amino acid sequences. ABG00010-ABG30377 represent novel human					
XX	diagnostic amino acid sequences of the invention.					
XX	Note: The sequence data for this patent did not appear in the printed					
XX	specification, but was obtained in electronic format directly from WIPO					
XX	at ftp.wipo.int/pub/published_pct_sequences.					
XX	Sequence 137 AA;					

	Query Match	57.9%	Score 66:	DB 22:	Length 137:
	Best Local Similarity	65.0%	Pred. No.	0.0094:	
	Matches 13:	Conservative 1:	Mismatches 6:	Indels 0:	Gaps 0:
OY	2 NSGDYKEDYGFAHITRADML 21 1 - - - - -				
Db	35 NQAKVKEHYGVSHIRHMDL 54				
RESULT 5					
ID	ABG04294 standard; Protein: 165 AA.				
XX	ABG04294;				
AC					
XX					
DT	13-FEB-2002 (first entry)				
XX					
DE	Novel human diagnostic protein #4285.				
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;				
RW	food supplement; medical imaging; diagnostic; genetic disorder.				
XX					
OS	Homo sapiens.				
PN	WO200175067-A2.				
PD	11-OCT-2001.				
PF	30-MAR-2001; 2001WO-US08631.				
XX					
PR	31-MAR-2000; 2000US-0540217.				
XX	23-AUG-2000; 2000US-0649167.				
PA	(HXSE-) HXSEQ INC.				
XX					
PI	Dmanac RT, Liu C, Tang YT;				
DR	WP1: 2001-639362/73.				
DR	N-PSDB: AAS68481.				
PT	New isolated polynucleotide and encoded polypeptides, useful in				
PT	diagnostics, forensics, gene mapping, identification of mutations				
PT	responsible for genetic disorders or other traits and to assess				
XX	biodiversity				
PS	Claim 20: SEQ ID NO 34653; 103pp; English.				
CC	The invention relates to isolated polynucleotide (I) and				
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,				
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome				
CC	and gene mapping, and in recombinant production of (II). The				
CC	polynucleotides are also used in diagnostics as expressed sequence tags				
CC	for identifying expressed genes. (I) is useful in gene therapy techniques				
CC	to restore normal activity of (II) or to treat disease states involving				
CC	(II). (II) is useful for generating antibodies against it, detecting or				
CC	quantitating a polypeptide in tissue, as molecular weight markers and as				
CC	a food supplement. (II) and its binding partners are useful in medical				
CC	imaging of sites expressing (II). (I) and (II) are useful for treating				
CC	disorders involving aberrant protein expression or biological activity.				
CC	The polypeptide and polynucleotide sequences have applications in				
CC	diagnostics, forensics, gene mapping, identification of mutations				
CC	and to produce other types of data and products dependent on DNA and				
CC	amino acid sequences. ABG00010-ABG03077 represent novel human				
CC	diagnostic amino acid sequences of the invention.				
CC	Note: The sequence data for this patent did not appear in the printed				
CC	specification, but was obtained in electronic format directly from Wipro				
CC	at ftp.wipo.int/pub/published_pct_sequences.				
SQ	Sequence 165 AA;				
Query Match	57.9%.	Score 66:	DB 22:	Length 165:	
Best Local Similarity	65.0%.	Pred. No.	0.0111:		

Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 NSGDYKEDYGFAMITRADML 21
 | ||| || : |||| |||
 Db 35 NQAKKETYGVSHITRDML 54

RESULT 6
 ABG04304
 ID ABG04304 standard; Protein: 175 AA.

AC ABG04304;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #4295.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS68491.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

PS Claim 20; SEQ ID No 34663; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 175 AA;

Query Match 57.9%; Score 66; DB 22; Length 175;

Best Local Similarity 65.0%; Pred. No. 0.012; Mismatches 6; Indels 0; Gaps 0;

OY 2 NSGDYKEDYGFAMITRADML 21
 | ||| || : |||| |||
 Db 35 NQAKKETYGVSHITRDML 54

RESULT 7
 ABG26186
 ID ABG26186 standard; Protein: 194 AA.

AC ABG26186;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #26177.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS90373.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

PS Claim 20; SEQ ID No 56545; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 194 AA;

Query Match 57.9%; Score 66; DB 22; Length 194;

Best Local Similarity 65.0%; Pred. No. 0.014; Mismatches 6; Indels 0; Gaps 0;

Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 NSGDYKEDYGFAMITRADML 21
 | ||| || : |||| |||

Db 134 NOKAYKETYGVSHTRHMDL 153

RESULT 8
ABG04299

ID ABG04299 standard; Protein; 231 AA.
XX
AC ABG04299;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #4290.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS68486.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 34658; 103pp; English.
XX

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences..
XX

SQL Sequence 231 AA:

Query Match 57.9%; Score 66; DB 22; Length 231;
Best Local Similarity 65.0%; Pred. No. 0.017;
Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 NSGDYKEDYGFAHTRADM 21
| ||| ||| : ||| |||
1 35 NOKAYKETYGVSHTRHMDL 54

Db

RESULT 9
ABG21816

ID ABG21816 standard; Protein; 241 AA.
XX
AC ABG21816;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21807.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS68003.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 52175; 103pp; English.
XX

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences..
XX

SQL Sequence 241 AA:

Query Match 57.9%; Score 66; DB 22; Length 241;
Best Local Similarity 65.0%; Pred. No. 0.017;
Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 NSGDYKEDYGFAHTRADM 21
| ||| ||| : ||| |||
1 35 NOKAYKETYGVSHTRHMDL 54

Db

RESULT 10

ABG04290
ID ABG04290 standard; Protein; 242 AA.
XX
AC ABG04290:
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #4281.
XX
DE Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSF-) HYSFO INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PI WPI; 2001-639362/73.
XX
DR N-PSDB; AAS68477.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20: SEQ ID No 34649; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 242 AA:
Query Match 57.9%; Score 66; DB 22; Length 242;
Best Local Similarity 65.0%; Pred. No. 0.017;
Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 NSGDYKEDYGFAHTRADM 21
I III I : III I I I I
DB 35 NOKAYKETYGSHIRHDM 54

RESULT 11
ABG04297
ID ABG04297 standard; Protein; 256 AA.

XX
AC ABG04297;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #4288.
XX
DE Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSF-) HYSFO INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PI WPI; 2001-639362/73.
XX
DR N-PSDB; AAS68484.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20: SEQ ID No 34656; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 256 AA:
Query Match 57.9%; Score 66; DB 22; Length 256;
Best Local Similarity 65.0%; Pred. No. 0.019;
Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 NSGDYKEDYGFAHTRADM 21
I III I : III I I I I
DB 35 NOKAYKETYGSHIRHDM 54

RESULT 12
AAU30761
ID AAU30761 standard; Protein; 266 AA.
XX
AC AAU30761;

[illegible]

OS	Homo sapiens.
XX	
PN	WO200175067-A2.
PD	
PP	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmacac RT, Liu C, Tang YT;
XX	
DR	WPI: 2001-639362/73.
XX	
N-PSDB:	AAS9376.
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity
XX	
PS	Claim 20; SEQ ID NO 56548; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from Wipo
CC	at ftp.wipo.int/pub/published_pct_sequences.
CC	
SQ	Sequence 267 AA:
Query Match	57.9%; Score 66; DB 22; Length 267;
Best Local Similarity	65.0%; Pred. No. 0.019;
Matches 13; Conservative	1; Mismatches 6; Indels 0; Gaps 0;
OY	2 NSGDYKEDYGFAHTRADM.L 21
	I I I I I I : I I I I I I I I
Db	157 NOKAYKETVGVSHIRHDM.L 176
RESULT 14	
ABGI0291	
ID	ABGI0291 standard; Protein; 268 AA.
XX	
AC	ABGI0291;
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #10282.
XX	
KM	Human; chromosome mapping; gene mapping; forensic;
KM	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	

```

XX 11-OCT-2001.
PD
PR 30-MAR-2001; 2001MO-US08631.
PR
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PR
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR
DR N-PSDB; AAS68511.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID NO 34683; 103pp: English.
PS
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC amino acid sequences.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 276 AA:
SQ
Query Match 57.9%; Score 66; DB 22; Length 276;
Best Local Similarity 65.0%; Pred. NO. 0.02;
Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0.
QY 2 NSGDYKEDYGFNAHITRADML 21
I IIII IIII IIII
DB 41 NOKAVKETYGVSHITRDHML 60

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:55:48 ; Search time 3.19672 Seconds
(without alignments)
500.202 Million cell updates/sec

Title: US-09-986-682B-2

Perfect score: 65

Sequence: 1 SVFDGCDTVYQ 12

Scoring table:

BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

A.Geneseq_101002:*
1: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	100.0	12	19	AAW48308
2	65	100.0	487	19	AAW48306
3	40	61.5	418	22	ABBS62267
4	39	60.0	185	17	AAW03548
5	39	60.0	608	23	ABBS7303
6	38.5	59.2	815	23	ABP35606
7	38	58.5	295	23	ABBS95607
8	38	58.5	295	23	ABBS5001
9	38	58.5	477	22	ABBS8663
10	38	58.5	558	13	AAW26843

11	38	58.5	582	23	ABBS1896
12	38	58.5	1192	22	ABBS3388
13	37.5	57.7	204	22	ABBS3290
14	37.5	57.7	1310	22	ABBS1027
15	37	56.9	103	23	ABBS4181
16	37	56.9	184	22	ABBS6957
17	37	56.9	185	22	ABBS9234
18	37	56.9	185	22	ABBS6958
19	37	56.9	190	20	AAV13480
20	37	56.9	238	22	ABBS2475
21	37	56.9	350	22	AAW60192
22	37	56.9	403	23	ABBS30257
23	37	56.9	408	21	ABBS1898
24	37	56.9	416	23	ABBS27335
25	37	56.9	587	22	AAU41035
26	37	56.9	597	22	ABBS60160
27	37	56.9	629	23	ABBS91126
28	37	56.9	769	23	ABBS91670
29	37	56.9	811	22	AAU36282
30	37	56.9	899	22	ABBS5488
31	37	56.9	899	22	ABBS5489
32	37	56.9	960	23	ABBS9642
33	37	56.9	1018	15	ABBS63759
34	37	56.9	1018	17	AAW87028
35	37	56.9	1018	18	AAW06485
36	37	56.9	1286	23	ABBS1476
37	37	56.9	2076	22	AAU34319
38	37	56.9	2186	22	AAU37320
39	36	55.4	16	22	AAW98607
40	36	55.4	37	19	AAW75153
41	36	55.4	63	20	AAW11775
42	36	55.4	113	22	AAE07005
43	36	55.4	135	22	ABBS8764
44	36	55.4	146	22	AAW64662
45	36	55.4	210	19	AAW39176

ALIGNMENTS

RESULT 1	
AAW48308	
ID	AAW48308 standard; Protein: 12 AA.
XX	
AC	AAW48308;
DT	04-JUN-1998 (first entry)
XX	
DE	Bacillus sp. V230 beta-fructofuranosidase internal fragment.
XX	
KW	Beta-fructofuranosidase; xylosyl fructoside; isomaltosyl fructoside;
KW	lactosucrose; fructosyltrehalose; antidiabetic; growth promoter;
KW	Bifidobacteria; mineral-absorbing; promoting activity; taste; texture;
XX	cosmetic; pharmaceutical; sweetener.
OS	Bacillus sp.
XX	
PN	EP812915-A2.
PD	17-DEC-1997.
XX	
PF	10-JUN-1997; 97EP-0304032.
XX	
PR	10-JUN-1996; 96JP-0170630.
PA	(HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
PI	Chaen H. Kubota M. Tsusaki K;
XX	
DR	WPI; 1998-034976/04.
XX	
PT	Bacillus beta-fructofuranosidase enzyme - useful for producing
PT	fructo:furanosylated saccharide(s) or alcohol(s)

XX Claim 1: Page 19; 32pp; English.
PS
XX
CC This is an internal fragment of beta-fructofuranosidase from
CC Bacillus sp. V230, an enzyme that has an optimum temperature of 50 deg. C
CC at pH 6. This enzyme catalyses fructofuranosyl transfer reactions from a
CC fructofuranosyl donor to a fructofuranosyl acceptor where the donor is
CC sucrose, raffinose or erlose and the acceptor is selected from alcohols,
CC 'sugar alcohols and saccharides having no beta-fructofuranosidic linkages,
CC especially where the reaction product is xylosyl fructoside, erlose,
CC isomaltosyl fructoside, lactosucrose or fructosyltrehalose.
CC Such reaction products have a satisfactory taste and sweetness, a
CC moderate viscosity and humectancy, an effective anticariogenic activity,
CC growth promoting activity for Bifidobacteria, mineral-absorption
CC promoting activity. The products can be used to improve the tastes and
CC textures of food products, cosmetics and pharmaceuticals and are useful
CC as sweeteners in the food and pharmaceutical industries.
CC
SQ Sequence 12 AA;
Query Match 100.0%; Score 65; DB 19; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SVFDGSDGTIVQ 12
DB 1 SVFDGSDGTIVQ 12
RESULT 2
AAW48306
ID AAW48306 standard; Protein; 487 AA.
XX
AC AAW48306;
XX
DT 04-JUN-1998 (first entry)
XX
DE Bacillus sp. V230 beta-fructofuranosidase.
XX
DE Beta-fructofuranosidase; xylosyl fructoside; isomaltosyl fructoside;
KW lacosucrose; fructosyltrehalose; anticariogenic; growth promoter;
KW Bifidobacteria; mineral-absorption promoting activity; taste; texture;
KW cosmetic; pharmaceutical; sweetener.
XX
OS Bacillus sp.
XX
XX
XX Key Location/Qualifiers
FH Peptide 1..32
FT /label= signal_peptide
FT 33..487
FT Protein /label= beta-fructofuranosidase
FT /note= "Claim 4"
XX
XX EP812915-A2.
XX
XX
XX PD 17-DEC-1997.
XX
XX PF 10-JUN-1997; 97EP-0304032.
XX
XX PR 10-JUN-1996; 96JP-0170630.
XX
XX (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
XX
XX Chaeen H, Kubota M, Tsusaki K;
XX
XX DR MPI: 1998-034976/04.
XX N-PSDB; AAV17621.
XX
XX Bacillus beta-fructofuranosidase enzyme - useful for producing
PT fructo:furanosylated saccharide(s) or alcohol(s)
PS
XX Claim 1; Pages 23-25; 32pp; English.

CC This is beta-fructofuranosidase from Bacillus sp. V230, an enzyme that
CC has an optimum temperature of 50 deg. C at pH 6. It catalyses
CC fructofuranosyl transfer reactions from a fructofuranosyl donor to a
CC fructofuranosyl acceptor where the donor is sucrose, raffinose or erlose
CC and the acceptor is selected from alcohols, sugar alcohols and
CC saccharides having no beta-fructofuranosidic linkages,
CC especially where the reaction product is xylosyl fructoside, erlose,
CC isomaltosyl fructoside, lactosucrose or fructosyltrehalose.
CC Such reaction products have a satisfactory taste and sweetness, a
CC moderate viscosity and humectancy, an effective anticariogenic activity,
CC growth promoting activity for Bifidobacteria, mineral-absorption
CC promoting activity. The products can be used to improve the tastes and
CC textures of food products, cosmetics and pharmaceuticals and are useful
CC as sweeteners in the food and pharmaceutical industries.
CC
SQ Sequence 487 AA;
Query Match 100.0%; Score 65; DB 19; Length 487;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SVFDGSDGTIVQ 12
DB 233 SVFDGSDGTIVQ 244
RESULT 3
ABB62267
ID ABB62267 standard; Protein; 418 AA.
XX
XX ABB62267;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13593.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX PD 23-MAR-2001; 2001WO-US09231.
XX
XX PF 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX DR MPI: 2001-656860/75.
XX N-PSDB; ABL06370.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions
XX
XX Disclosure; SEQ ID NO 13593; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (AB101840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 418 AA:

Query Match 61.5%; Score 40; DB 22; Length 413;

Best Local Similarity 70.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 VFDDGDTGVY 11

DB 13 LFDDGDMNVY 22

RESULT 4

ID AAM03548 standard; Protein: 185 AA.

XX AAM03548;

DT 21-APR-1997 (first entry)

DE ORF-2 protein sequence from BamHI fragment of HVT.

XX Herpes virus of turkey; open reading frame; ORF; homology; vector;
 KW avian herpes virus; recombinant viral vaccine; intergenic region; IBDV;
 KW cytomegalovirus immediate early promoter; UL55 gene; repeat region; ILTV;
 KW antigen; infectious bursal disease virus; Marek's disease virus; MDV;
 KW infectious laryngotracheitis virus; avian anaemia virus; vaccination;
 KW infectious bronchitis virus; IBV; poultry; Gumboro disease;
 KW Newcastle disease.

XX Herpesvirus of turkey.

OS EP119864-A2.

XX 03-JUL-1996.

XX 28-DEC-1995; 95EP-0402970.

XX 30-DEC-1994; 94FR-0016017.

XX (INMR) RHONE MERIEUX SA.

XX Audonnet JCF, Bublout MJM, Dartell RJ, Duinat CV;

PI Laplace ELF, Riviere MAE;

XX WPI: 1996-364150/37.

DR N-PSDB: AAT39309.

XX Live recombinant avian vaccine - comprises herpes virus as vector
 PT and having sequence encoding antigenic polypeptide inserted between
 PT UL55 gene and repeat region

XX Example 4; Fig 1: 50pp; French.

XX This is the amino acid sequence of the protein encoded by the open
 CC reading frame 2 (ORF-2) from the BamHI I fragment from the herpes virus
 CC of turkeys (HVT). The protein has homology to the 185 amino acid protein
 CC encoded by the EHV-1 gene 3 ORF. The DNA fragment was isolated from HVT
 CC strain FC126 (J. Gen. Virol., 70:1789-1804 (1989)) and is used as a basis
 CC for generating live recombinant viral vaccines using the herpes virus DNA
 CC sequences as vectors. The fragment contains 3 intergenic regions into
 CC which genes encoding heterologous proteins to be expressed can be
 CC inserted, preferably under control of the cytomegalovirus immediate early
 CC (CMV-IE) promoter. The heterologous proteins can be used to vaccinate
 CC poultry against Gumboro disease (caused by IBDV), Newcastle disease,
 CC Marek's disease, infectious bronchitis, infectious laryngotracheitis and
 CC avian anaemia.

XX Sequence 185 AA:

Query Match 60.0%; Score 39; DB 17; Length 185;

Best Local Similarity 60.0%; Pred. No. 78;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 VFDDGDTGVY 11

DB 74 IFEGDGSVY 83

RESULT 5

ID ABB97303 standard; Protein: 608 AA.

XX ABB97303;

XX 27-JUN-2002 (first entry)

DE Novel human protein SEQ ID NO: 571.

XX Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
 KW antiinfectivity; cerebroprotective; cytosolic; rheumatic; gene therapy;
 KW neuroprotective; antiparkinsonian; protein therapy; EST;
 KW expressed sequence tag.

XX Homo sapiens.

XX WO200222660-A2.

XX 21-MAR-2002.

XX 10-SEP-2001; 2001WO-US26015.

XX 11-SEP-2000; 2000US-0659671.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI: 2002-292408/33.

DR N-PSDB: ABB32489.

XX An isolated polynucleotide for treating diseases associated with its
 PT encoded polypeptide such as cancer and multiple sclerosis -

XX Example 2; SEQ ID NO 571; 50pp; English.

XX The present invention provides the protein and coding sequences of 444
 CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to regulate
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a protein of the invention.

XX Sequence 608 AA:

Query Match 60.0%; Score 39; DB 23; Length 608;

Best Local Similarity 54.5%; Pred. No. 2.9e+02;

Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 VFDDGDTGVYQ 12

DB 310 LFEGGGSVFQ 320

RESULT 6

ID ABB35606 standard; Protein: 815 AA.

XX ABB35606;

XX ABB35606;

DT		24-JUL-2002	(first entry)
DE		Fungal ZBC protein sequence #32.	
KW		Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;	
KX		antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastatin;	
KM		mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;	
KV		angio genesis inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;	
KW		fungal toxin; cell surface receptor; plant growth regulator; pigment;	
KX		insecticide; antineoplastic.	
OS		Unidentified.	
PN		WO200224865-A2.	
PD		28-MAR-2002.	
PX		19-SEP-2001; 2001WO-US29288.	
PR		19-SEP-2000; 2000US-233564P.	
PA		(MICR-) MICROBIA INC.	
PI		Holtzman D, Madden K, Maxon M, Sherman A:	
DH		WPI; 2002-352005/38.	
DR		N-Psdb; ABN79795.	
PT		New method for improving the production of a secondary metabolite e.g.	
PT		antineoplastic agent, ergot alkaloid from a fungus involves modulation	
PT		of the expression of at least one zinc binuclear cluster protein gene	
PS		Disclosure; SEQ ID 76; 49pp + sequence listing; English.	
XX			
CC		The invention relates to improving the production of a secondary	
CC		metabolite by a fungus. This involves modulating the expression of at	
CC		least one ZBC (zinc binuclear cluster protein) gene in a manner to	
CC		improve the yield of the secondary metabolite. Methods of the invention	
CC		may be used for improving the production of the secondary metabolite e.g.	
CC		antibacterial (such as beta-lactam), an anti-hypercholesterolaemic (such	
CC		as lovastatin or mevastatin), an immunosuppressant (such as cyclosporin A),	
CC		an ergot alkaloid (such as ergotamine), an angiogenesis inhibitor (such	
CC		as ovalicin), a glucan synthase inhibitor, gliotoxin family of compounds,	
CC		a fungal toxin, a modulator of cell surface receptor signalling, a plant	
CC		growth regulator, a pigment, an insecticide, or an antineoplastic	
CC		compound. The method results in a decrease in fermentor run-time, a	
CC		decrease in the size of the fermentor required for the production of	
CC		equivalent amounts of the secondary metabolite, or a decrease in the	
CC		biomass required for the production, which translates into decreased	
CC		waste that must be handled in downstream processing. The sequences given	
CC		in records ABP35575-ABP35722 represent ZBC proteins.	
CC		Note: The sequence data for this patent did not form part of the printed	
CC		specification, but was obtained directly from WIPO at	
CC		ftp.wipo.int/pub/published_pct_sequences.	
XX			
Sequence	815 AA:		
OY			
DB			
Query Match	59.2%;	Score 38.5; DB 23; Length 815;	
Best Local Similarity	69.2%;	Pred. No. 4.8e+02;	
Matches 9; Conservative 1;	Mismatches 2;	Indels 1; Gaps 1;	
1 SVFDGGDGTV-YQ 12			
:			
680 SVFDGNIGTIPYQ 692			
RESULT 7			
ID ABB95607			
AC ABB95607 standard; Protein: 295 AA.			
XX ABB95607;			
XX			

DT		19-JUL-2002	(first entry)
XX			
DE			Human angiogenesis related protein PRO28631 SEQ ID NO: 370.
XX			
KW			Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
KV			atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder
KX			candidati; cytoslastic; antiangiogenic; hypotensive; vulnerary;
KW			antiarteriosclerotic.
XX			
OS	Homo sapiens.		
PN	MO200208284-A2.		
XX			
PD	31-JAN-2002.		
XX			
PF	09-JUL-2001; 2001WO-US21735.		
XX			
20-JUL-2000;	2000US-219556P.		
PR	25-JUL-2000; 2000US-22062AP.		
PR	25-JUL-2000; 2000US-22066AP.		
PR	28-JUL-2000; 2000WO-US20710.		
PR	02-AUG-2000; 2000US-222695P.		
PR	17-AUG-2000; 2000US-064365T.		
PR	23-AUG-2000; 2000WO-US23522.		
PR	24-AUG-2000; 2000WO-US23328.		
PR	07-SEP-2000; 2000US-230978P.		
PR	15-SEP-2000; 2000US-000000P.		
PR	18-SEP-2000; 2000US-0664610.		
PR	18-SEP-2000; 2000US-0665350.		
PR	24-OCT-2000; 2000US-242922P.		
PR	08-NOV-2000; 2000US-070923P.		
PR	08-NOV-2000; 2000WO-US30952.		
PR	10-NOV-2000; 2000WO-US30873.		
PR	01-DEC-2000; 2000WO-US32678.		
PR	20-DEC-2000; 2000US-0747259.		
PR	20-DEC-2000; 2000WO-US34956.		
PR	22-JAN-2001; 2001US-0767609.		
PR	28-FEB-2001; 2001US-0796498.		
PR	28-FEB-2001; 2001WO-US06520.		
PR	01-MAR-2001; 2001WO-US06666.		
PR	09-MAR-2001; 2001US-0802706.		
PR	14-MAR-2001; 2001US-0808689.		
PR	22-MAR-2001; 2001US-0816744.		
PR	05-APR-2001; 2001US-0828366.		
PR	10-MAY-2001; 2001US-0854208.		
PR	10-MAY-2001; 2001US-0854280.		
PR	25-MAY-2001; 2001US-0866028.		
PR	25-MAY-2001; 2001US-0866034.		
PR	25-MAY-2001; 2001WO-US17092.		
PR	30-MAY-2001; 2001US-0870574.		
PR	30-MAY-2001; 2001WO-US17443.		
PR	01-JUN-2001; 2001WO-US17800.		
PR	20-JUN-2001; 2001WO-US19692.		
PR	28-JUN-2001; 2001WO-US00000.		
XX			
PA	(GETH) GENENTECH INC.		
PA	(BAKE/) BAKER K P.		
PA	(FERR/) FERRARA N.		
PA	(GERB/) GERBER H.		
PA	(GERR/) GERRITSEN M E.		
PA	(GODD/) GODDARD A.		
PA	(GODO/) GODOWSKI P J.		
PA	(GURO/) GURNEY A L.		
PA	(HILL/) HILLAN K J.		
PA	(MARS/) MARSTERS S A.		
PA	(PAN/) PAN J.		
PA	(PRON/) PRONI N F.		
PA	(STEP/) STEPHAN J F.		
PA	(MATA/) WATANABE C K.		
PA	(WILL/) WILLIAMS P M.		
PA	(WOOD/) WOOD W I.		
XX			
I1	Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;		

PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
 PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
 XX MPI: 2002-171999/22.
 DR N-PSDB: ABL95745.
 XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,
 PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
 PT infarction), endothelial or angiogenic disorders in a mammal -
 XX
 PS Claim 11; Fig 370; 567pp; English.
 CC The present invention provides the protein and coding sequences of human
 CC PRO proteins. These are useful for treating or diagnosing a
 CC cardiovascular, endothelial or angiogenic disorder, including cardiac
 CC hypertrophy, trauma, cancer, age-related macular degeneration,
 CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
 CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
 CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
 CC healing. The present sequence is a PRO protein of the invention.
 CC
 XX Sequence 295 AA;
 SQ

Query Match 58.5%; Score 38; DB 23; Length 295;
 Best Local Similarity 75.0%; Pred. No. 1.9e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 FDGGDGTV 10
 |||||
 Db 58 FDGGDATT 65

RESULT 8
 ABB85001
 ID ABB85001 standard; Protein: 295 AA.

AC ABB85001;
 DT 16-MAY-2002 (first entry)

DE Human PRO28631 protein sequence SEQ ID NO:370.

XX Human; angiogenesis; cardiant; cytosolic; antiangiogenic; hypotensive;
 KW vulnerability; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
 KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
 KW angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
 KW age-related macular degeneration; arterial restenosis; angina;
 KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
 KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
 KW wound healing; chromosome mapping; gene mapping.

XX Homo sapiens.

XX WO200200690-A2.

XX 03-JAN-2002.

PF 20-JUN-2001; 2001WO-US19692.

XX 23-JUN-2000; 2000US-213637P.
 PR 20-JUL-2000; 2000US-219556P.
 PR 25-JUL-2000; 2000US-220624P.
 PR 28-JUL-2000; 2000US-220664P.
 PR 02-AUG-2000; 2000US-222695P.
 PR 17-AUG-2000; 2000US-0643657.
 PR 23-AUG-2000; 2000WO-US23322.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 07-SEP-2000; 2000US-230978P.
 PR 18-SEP-2000; 2000US-0664610.
 PR 24-OCT-2000; 2000US-242922P.
 PR 08-NOV-2000; 2000US-0709238.

PR 08-NOV-2000; 2000WO-US30952.
 PR 10-NOV-2000; 2000WO-US30873.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000US-0747259.
 PR 20-DEC-2000; 2000WO-US34956.
 PR 22-JAN-2001; 2001US-0767609.
 PR 28-FEB-2001; 2001US-0796498.
 PR 28-FEB-2001; 2001WO-US06520.
 PR 01-MAR-2001; 2001WO-US06666.
 PR 09-MAR-2001; 2001US-0802706.
 PR 14-MAR-2001; 2001US-0808689.
 PR 22-MAR-2001; 2001US-0816744.
 PR 05-APR-2001; 2001US-0828366.
 PR 10-MAY-2001; 2001US-0854208.
 PR 10-MAY-2001; 2001US-0854280.
 PR 25-MAY-2001; 2001US-0866028.
 PR 25-MAY-2001; 2001US-0866034.
 PR 25-MAY-2001; 2001US-0870574.
 PR 30-MAY-2001; 2001WO-US17092.
 PR 30-MAY-2001; 2001WO-US17443.
 PR 01-JUN-2001; 2001WO-US17800.

XX (GETH) GENENTECH INC.

XX Baker KP, Ferrara N, Gerber H, Gertlisen ME, Goddard A;
 PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
 PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
 XX MPI: 2002-090516/12.

DR N-PSDB: ABL88256.
 XX

PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
 PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
 PT infarction), endothelial or angiogenic disorders in a mammal -
 XX
 PS Claim 11; Fig 370; 565pp; English.

XX ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to
 CC ABB85003. The PRO proteins and polynucleotides have cardiant, cytosolic,
 CC antiangiogenic, hypotensive, vulnerability, antiarteriosclerotic
 CC activities, and can be used in gene therapy. The PRO polynucleotides,
 CC proteins, agonists and antagonists are useful for treating or diagnosing
 CC a cardiovascular, endothelial or angiogenic disorder in a mammal,
 CC e.g. cardiac hypertrophy, trauma, cancer, age-related macular
 CC degeneration, atherosclerosis, hypertension, arterial restenosis,
 CC rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis,
 CC lymphangitis, tumour angiogenesis (such as breast carcinoma and liver
 CC carcinoma) and wound healing. The PRO polynucleotides have applications
 CC in molecular biology, including use as hybridisation probes, and in
 CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
 CC probes used in the exemplification of the present invention.

XX Sequence 295 AA;

Query Match 58.5%; Score 38; DB 23; Length 295;
 Best Local Similarity 75.0%; Pred. No. 1.9e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 FDGGDGTV 10
 |||||
 Db 58 FDGGDATT 65

RESULT 9
 ABB58863
 ID ABB58863 standard; Protein: 477 AA.

XX ABB58863;
 AC

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 3381.
 XX

KW Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR N-PSDB: ABL02966.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 3381; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB57737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SO Sequence 477 AA;
 Query Match 58.5%; Score 38; DB 22; Length 477;
 Best Local Similarity 72.7%; Pred. No. 3.3e+02;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 SVFDGSGDTVY 11
 I:| | | | | | |
 Db 189 SMFLGFDGTVY 199
 RESULT 10
 AAR26843
 ID AAR26843 standard; Protein: 558 AA.
 XX
 AC AAR26843;
 XX
 DT 11-FEB-1993 (first entry)
 XX
 DE GDP dissociation stimulatory protein.
 XX
 KM Guanosine 5'-diphosphate; mass production.
 XX
 OS Not known.
 OS
 PN JP04211700-A.
 XX
 PD 03-AUG-1992.
 XX
 PF 08-FEB-1991; 91JP-0018105.
 XX
 PR 24-MAY-1990; 90JP-0134480.
 XX
 PA (MITU) MITSUBISHI KASEI CORP.

XX
 DR WPI: 1992-305013/37.
 DR N-PSDB: AAQ28232.
 XX
 PT High mol. wt. mass producible protein - accelerates dissociation
 PT of guanosine 5'-di-phosphate from complex of GDP and GDP protein
 PT binding type low mol guanosine 5'-tri-phosphate
 XX
 PS Claim 1; Fig 1; 17pp; Japanese.
 XX
 CC The sequence is that of a protein which accelerates dissociation of
 CC guanosine 5'-di-phosphate (GDP) from a complex of GDP and GDP protein
 CC binding type low molecular guanosine 5'-triphosphate (GTP) binding
 CC protein. It may be mass produced using genetic engineering techniques.
 XX
 SO Sequence 558 AA;
 Query Match 58.5%; Score 38; DB 13; Length 558;
 Best Local Similarity 54.5%; Pred. No. 3.9e+02;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 2 VFDGSGDTVYQ 12
 I:| | | | | | |
 Db 260 LFEKGKGNVFQ 270
 RESULT 11
 ABG61896
 ID ABG61896 standard; Protein: 582 AA.
 XX
 AC ABG61896;
 XX
 DT 15-AUG-2002 (first entry)
 XX
 DE Prostate cancer-associated protein #97.
 XX
 KM Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.
 XX
 OS Mammalia.
 OS
 PN WO200230268-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US32045.
 XX
 PR 13-OCT-2000; 2000US-0687576.
 PR 08-DEC-2000; 2000US-0733288.
 PR 08-DEC-2000; 2000US-0733742.
 PR 24-JAN-2001; 2001US-263957P.
 PR 16-MAR-2001; 2001US-276791P.
 PR 16-MAR-2001; 2001US-276888P.
 PR 06-APR-2001; 2001US-281922P.
 PR 24-APR-2001; 2001US-286214P.
 PR 30-APR-2001; 2001US-0847046.
 PR 04-MAY-2001; 2001US-288589P.
 XX
 PA (EOSB-) EOS BIOTECHNOLOGY INC.
 XX
 PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezl P;
 XX
 DR WPI: 2002-471335/50.
 DR N-PSDB: ABK92212.
 XX
 PT Detecting a prostate cancer-associated transcript in a cell in a
 PT patient, useful for diagnosing prostate cancer (PC) or screening
 PT modulators of PC, by determining if prostate cancer-associated genes
 PT are expressed in a prostate tissue
 XX
 PS Claim 27; Page 383; 436pp; English.
 XX
 CC The present invention relates to methods of detecting a prostate
 CC cancer-associated transcript in a cell from a patient. The method

CC comprises contacting a biological sample from the patient with
CC prostate cancer-associated polynucleotides (designated PC genes) that
CC selectively hybridize to a sequence that is at least 80% identical
CC to them. The prostate cancer-associated polynucleotide sequences
CC are differentially expressed in prostate tumour tissue or in
CC prostate cancer and are derived from the tissues of various
CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).
CC The methods of the invention are useful for diagnosing and treating
CC prostate cancer in mammals. The prostate cancer-associated genes are
CC useful for diagnosing or treating prostate cancer, as well as for
CC identifying modulators of prostate cancer or agents that inhibit
CC prostate cancer. The nucleic acid sequences are particularly useful
CC in gene therapy, as a vaccine or in antisense applications.
CC ABG61800-ABG61944 represent prostate cancer-associated proteins.

SO Sequence 582 AA:

Query Match 58.5%; Score 38; DB 23; Length 582;
Best Local Similarity 75.0%; Pred. No. 4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 FDGGDGTV 10
| | | | |
Db 278 FDGGDXTI 285

RESULT 12

ABG93388
ID ABG93388 standard; Protein; 1192 AA.

AC ABB93388;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 2599.

KW Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN MO200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001MO-EP09892.

PR 28-AUG-2001; 2001MO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidler M;

DR WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -

PS Claim 5; SEQ ID NO 2599; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins
XX (ABG90790-ABG94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.

SO Sequence 1192 AA:

Query Match 58.5%; Score 38; DB 23; Length 1192;
Best Local Similarity 77.8%; Pred. No. 8.9e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 DGGDGTIVQ 12
| | | | |
Db 925 DGGGTIVTK 933

RESULT 13

ABG13290
ID ABG13290 standard; Protein; 204 AA.

AC ABG13290;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #13281.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN MO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001MO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS77477.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID NO 43649; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX characterizing a polypeptide in tissue, as molecular weight markers and as
XX a food supplement (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 204 AA:

Query Match 57.7%; Score 37.5; DB 22; Length 204;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;

Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;
QY 1 SVEPDGDTGYQ 12
|:|:| | | | |
Db 176 SIFD-GDCKTYQ 186

RESULT 14
ABG10127
ID ABG10127 standard; Protein: 1310 AA.
XX

AC ABG10127;
XX

DT 13-FEB-2002 (first entry)
XX

DE Novel human diagnostic protein #10118.
XX

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX

OS Homo sapiens.
XX

PN W0200175067-A2.
XX

PD 11-OCT-2001.
XX

PF 30-MAR-2001; 2001WO-US08631.
XX

PR 31-MAR-2000; 2000US-0540217.
XX

PR 23-AUG-2000; 2000US-0649167.
XX

PA (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YT;
XX

PI N-PSDB: AAS74314.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX

PS Claim 20; SEQ ID NO 40486; 103pp; English.
XX

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 1310 AA;
XX

Query Match 57.7%; Score 37.5; DB 22; Length 1310;
Best Local Similarity 66.7%; Pred. No. 1.2e+03;
Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 SVEPDGDTGYQ 12
|:|:| | | | |
Db 1282 SIFD-GDCKTYQ 1292

RESULT 15
ABP34181
ID ABP34181 standard; Protein: 103 AA.
XX

AC ABP34181;
XX

DT 08-JUL-2002 (first entry)
XX

DE Human glycoprotein-like ORF3154 protein, SEQ ID NO:6308.
XX

KW Human: ORF; open reading frame; ORF; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokine; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW cardiac; hypotensive; antihypertoid; antiinflammatory; immunomodulator;
KW dermatological; analgesic; virucide; antibacterial; fungicide.
XX

OS Homo sapiens.
XX

PN W0200190366-A2.
XX

PD 29-NOV-2001.
XX

PF 24-MAY-2001; 2001WO-US17076.
XX

PR 24-MAY-2000; 2000US-206690P.
XX

PA (CURA-) CURAGEN CORP.
XX

PI Leach MD, Shinkels RA;
XX

PI N-PSDB: ABN78207.
XX

PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation -
XX

PS Claim 10; Page 1826-1827; 2508pp; English.
XX

CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antiinfective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,

CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases.
XX

SQ Sequence 103 AA;

Query Match 56.9%; Score 37; DB 23; Length 103;
Best Local Similarity 77.8%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 DGGDGTVTYQ 12
|||||
Db 66 DGGDGVVSQ 74

Search completed: June 16, 2003, 12:06:46
Job time : 4.19672 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:01:28 ; Search time 3.02459 Seconds
(without alignments)
817.488 Million cell updates/sec

Title: US-09-986-682B-2
Perfect score: 65
Sequence: 1 SVFDGSDGTIVYQ 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	100.0	487	2	082854
2	59	90.8	489	16	097181
3	54	83.1	499	2	0925E5
4	45	69.2	800	2	09F8X1
5	44	67.7	125	16	08RI28
6	44	67.7	888	17	0977D4
7	43	66.2	330	16	09A934
8	42	64.6	926	5	020176
9	41	63.1	239	10	09F1C9
10	41	63.1	263	5	09BLL7
11	41	63.1	295	16	09Kf21
12	41	63.1	380	16	09ZBK8
13	41	63.1	2658	13	090WF0
14	40	61.5	225	16	086806
15	40	61.5	418	5	09VTD9
16	40	61.5	469	16	050499

17	39	60.0	108	10	09M904	09m9u4 arabidopsis
18	39	60.0	175	10	08RXS1	08rxs1 arabidopsis
19	39	60.0	185	12	09DGT5	09dgt5 melegrid h
20	39	60.0	212	10	09FTK1	09ftk1 oryza sativ
21	39	60.0	346	16	09RD67	09rd67 streptomyc
22	39	60.0	406	4	09B0W9	09b0w9 homo sapien
23	39	60.0	444	10	09XHP2	09xhp2 zea mays (m
24	39	60.0	523	8	09ZY25	09zy25 pedionomas
25	39	60.0	558	11	092105	092105 mus musculu
26	39	60.0	607	4	09NZAB	09nza8 homo sapien
27	39	60.0	607	4	09NYM2	09nym2 homo sapien
28	39	60.0	607	13	09B0X6	09b0x6 homo sapien
29	39	60.0	607	13	093614	093614 xenopus lae
30	39	60.0	1299	16	08XWT7	08xwt7 raltosnia s
31	39	60.0	1420	10	09XEJ3	09xel3 oryza sativ
32	39	60.0	2185	9	08W6J4	08w6j4 sinorhizobi
33	39	60.0	2454	5	08T2G3	08t2g3 dictyostell
34	39	60.0	4283	11	09ERV0	09erv0 ratius norv
35	39	60.0	4293	11	098852	098852 mus musculu
36	38.5	59.2	647	17	09HOM8	09hgm8 halobacteri
37	38.5	59.2	815	3	059744	059744 schizosacch
38	38	58.5	146	10	023043	023043 arabidopsis
39	38	58.5	214	7	09TNV6	09tnv6 oncorhynch
40	38	58.5	295	11	08QZV2	08qzv2 mus musculu
41	38	58.5	358	2	09X5K8	09x5k8 streptomyc
42	38	58.5	364	2	007862	007862 streptomyc
43	38	58.5	364	16	09RJY2	09rjy2 streptomyc
44	38	58.5	398	16	067179	067179 aquifex aeo
45	38	58.5	418	16	09WYC7	09wyc7 thermotoga

ALIGNMENTS

RESULT 1	ID	PRELIMINARY:	PRT:	487 AA.
AC	082854	082854		
DT	01-NOV-1998	(TREMBLrel. 08, Created)		
DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)		
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)		
DE	Beta-fructofuranosidase.			
OS	Bacillus sp.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;			
OC	Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1409;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-V230;			
RC	STRAIN-V230;			
RA	Kurimoto M.;			
RT	"Cloning and sequencing of beta-fructofuranosidase gene from Bacillus			
RT	sp. V230."			
SO	Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-V230;			
RC	STRAIN-V230;			
RA	Kurimoto M.;			
RT	"Cloning and sequencing of beta-fructofuranosidase gene from Bacillus			
RT	sp. V230."			
SO	Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AB010272; BAA32083.1;			
DR	InterPro; IPR003469; Glyco_hydro.68.			
DR	Pfam; PF02435; Glyco_hydro.68; 1.			
SO	SEQUENCE 487 AA; 53412 MW; 7F52AA7921824AE3 CRC64;			

Query Match	100.0%;	Score 65;	DB 2;	Length 487;
Best Local Similarity	100.0%;	Pred. No. 0.017;		
Matches 12;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	1 SVFDGSDGTIVYQ 12			
Db	233 SVFDGSDGTIVYQ 244			

RESULT 2

097181 PRELIMINARY; PRT: 489 AA.
ID 097181;
AC 097181;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN CAC1772.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RA MEDLINE-21359325; PubMed-11466286;
RT Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
Gibson R., Lee H.M., Dupont J., Qiu D., Hiltl J., Wolf Y.T.,
Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007686; AAK79737.1;
DR InterPro: IPR003469; Glyco_hydro.68.
DR Pfam: PF02435; Glyco_hydro.68; 1;
KW Complete proteome.
SQ SEQUENCE 489 AA; 54553 MW; 9002CB36AF1D3CD8 CRC64;

Query Match 90.8%; Score 59; DB 16; Length 489;
Best Local Similarity 83.3%; Pred. No. 0.16;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SVFDGDDGTYYQ 12
DB 229 SVFDGDDGTYYQ 240
|||||:|||||

RESULT 3
0925E5 PRELIMINARY; PRT: 499 AA.
ID 0925E5;
AC 0925E5;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN Levanusucrase.
OS Paenibacillus polymyxa (Bacillus polymyxa).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=1406;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-CF43;
RA MEDLINE-21042000; PubMed-11200435;
RA Bezate S., Aymeric S., Chamber R., Czarnez S., Berge O., Heulin T.;
RT "Disruption of the Paenibacillus polymyxa levanusucrase gene impairs
its ability to aggregate soil in the wheat rhizosphere.";
RL Environ. Microbiol. 2:333-342(2000).
DR EMBL: AJ133737; CAB39327.1;
DR InterPro: IPR003469; Glyco_hydro.68.
DR Pfam: PF02435; Glyco_hydro.68; 1;
SQ SEQUENCE 499 AA; 55317 MW; E646CD986292336C CRC64;

Query Match 83.1%; Score 54; DB 2; Length 499;
Best Local Similarity 66.7%; Pred. No. 1.1;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SVFDGDDGTYYQ 12
DB 231 SVFDGDDGTYYQ 242
|||||:|||||

RESULT 4
09F8X1 PRELIMINARY; PRT: 800 AA.
ID 09F8X1;
AC 09F8X1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
GN CHBP.
OS Vibrio furnissii.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=29494;
[1]
RP SEQUENCE FROM N.A.
RC MEDLINE-20490751; PubMed-10913116;
RA Park J.K., Keyhani N.O., Roseman S.;
RT "Chitin catabolism in the marine bacterium Vibrio furnissii.
IDENTIFICATION, MOLECULAR CLONING, AND CHARACTERIZATION OF A N,N'-
DIACETYLCHITOSIDASE PHOSPHORYLASE.";
RT J. Biol. Chem. 275:33077-33083(2000).
DR EMBL: AF230379; AAG23740.1;
SQ SEQUENCE 800 AA; 89666 MW; 6775DB23C051122A CRC64;

Query Match 69.2%; Score 45; DB 2; Length 800;
Best Local Similarity 88.9%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 DGGDGTYYQ 12
DB 456 DGGDGTYYQ 464
|||||:|||||

RESULT 5
08R128 PRELIMINARY; PRT: 125 AA.
ID 08R128;
AC 08R128;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
GN FNI808.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 25586;
RA MEDLINE-21886394; PubMed-11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Gotsman E., Bernal A.,
RA Larsen N., D'Souza M., Malinas T., Pusch G., Haselkorn R.,
RA Fongstein M., Kyrides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
nucleatum strain ATCC 25586.";
RT J. Bacteriol. 184:2005-2018(2002).
DR EMBL: AE010483; AAL3907.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 125 AA; 13549 MW; 661F5D709B72ABC5 CRC64;

Query Match 67.7%; Score 44; DB 16; Length 125;
Best Local Similarity 72.7%; Pred. No. 10;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SVFDGDDGTYYQ 11
DB 11 SIDDGDDGTYYQ 21
|||||:|||||

RESULT 6
0977D4 PRELIMINARY; PRT: 888 AA.
ID 0977D4;
AC 0977D4;
SQ SEQUENCE 888 AA; 90774 MW; 661F5D709B72ABC5 CRC64;

DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Hypothetical protein ST0012.
 GN ST0012.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 10545 / 7;
 RX PubMed-11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 Sakane M., Baba S.-I., Anai A., Kosugi H., Hosoyama A., Fukui S.,
 Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
 Aoki K.-I., Masuda S., Yanagii M., Nishimura J., Yamagishi A.,
 Oshima T., Kikuchi H.;
 RA "Complete genome sequence of an aerobic thermocidophilic
 RT Crenarchaeon, Sulfolobus tokodaii strain7.";
 RL DNA Res. 8:123-140(2001).
 DR EMBL, AP000981; BAB64960.1; -
 DR InterPro: IPR002819; HD; -
 DR InterPro: IPR003607; ME_Plpase_HDC.
 DR Pfam: PF01966; HD; 1.
 DR SMART: SM00471; HDC; 1.
 DR TIGRFAMS: TIGR00277; HDIG; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 888 AA; 101756 MW; 008CC23F450C14F CRC64;

Query Match Best Local Similarity 67.7%; Score 44; DB 17; Length 888;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 VFDDGDTVYQ 12
 Db 1111111111
 15 VFDDGDKVYE 25

RESULT 7
 Q9A934 PRELIMINARY; PRT; 330 AA.
 AC Q9A934;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Hypothetical protein CC1160.
 GN CC1160.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 19089 / CB15;
 RX MEDLINE-21173698; PubMed-11259647;
 RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 Polocka I., Nelson W.C., Newton A.S., Stephens C., Phadke N.D., Ely B.,
 DeBoy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
 Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty K.,
 Uterback T., Tran K., Wolf A., Vamathavan J., Ermolaeva M., White O.,
 Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RA "Complete genome sequence of Caulobacter crescentus.";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 RL EMBL, AE005794; AAK3144.1; -
 DR TIGR: CC1160; -
 DR InterPro: IPR001206; DAGKC.
 DR Pfam: PF00781; DAGKC; 1.
 DR SMART: SM00046; DAGKC; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 330 AA; 35352 MW; 4112B72A9615E7C9 CRC64;

Query Match Best Local Similarity 66.2%; Score 43; DB 16; Length 330;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VFDDGDTV 10
 Db 1111111111
 80 VFDDGDTV 88

RESULT 8
 ID Q20176 PRELIMINARY; PRT; 926 AA.
 AC Q20176;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Hypothetical 105.3 kDa protein.
 GN F38E9.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE-99069613; PubMed-9851916;
 RA None.
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Gattung S., Wu X.;
 RT "The sequence of C. elegans cosmid F38E9.";
 RT Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RT Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 9 CUB DOMAINS.
 DR EMBL: U46668; AAA93348.2; -
 DR HSP, P00736; IAPQ.
 DR InterPro: IPR001506; Astacin.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01400; Astacin; 2.
 DR Pfam: PF00431; CUB; 8.
 DR Pfam: PF00008; EGF; 4.
 DR SMART: SM00042; CUB; 9.
 DR SMART: SM00181; EGF; 2.
 DR SMART: SM00179; EGF_Ca; 4.
 DR SMART: SM00235; Zmnc; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 4.
 DR PROSITE: PS01180; CUB; 8.
 DR PROSITE: PS01186; EGF_2; 4.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN.1.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation;
 KW Hypothetical protein; Repeat.
 SQ SEQUENCE 926 AA; 105287 MW; A7E6F04B28B77D00 CRC64;

Query Match Best Local Similarity 64.6%; Score 42; DB 5; Length 926;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SVFDDGDT 9

DB 751 SVYDGGDGS 759

RESULT 9

Q9FLC9 PRELIMINARY: PRT: 239 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Genomic DNA, chromosome 5, TAC clone:K18123 (A15905250/K18123_5).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=98344145; PubMed=9679202;
 RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
 RT features of the regions of 1,381,365 bp covered by twenty one
 RT physically assigned P1 and TAC clones."
 RL DNA Res. 5:131-145(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RT Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AB010692; BAB09968.1;
 DR EMBL; AY054291; AAL06949.1;
 SQ SEQUENCE 239 AA; 25863 MW; 483C313F537EE2CA CRC64;

Query Match 63.1%; Score 41; DB 10; Length 239;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 FDGGDGT 9

DB 166 FDGGDGT 172

RESULT 10

Q9BL7 PRELIMINARY: PRT: 263 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Fibroin L-chain.
 GN FIB-L.
 OS Dendrolimus spectabilis (pine moth).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 OC Bombycoidea; Lasiocampidae; Dendrolimus.
 NCBI_TaxID=155323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21167527; PubMed=11267905;
 RA Tanaka K., Mizuno S.;
 RT "Homologues of fibroin L-chain and P25 of Bombyx mori are present in
 RT Dendrolimus spectabilis and Papilio xuthus but not detectable in
 RT Antherea yamamai.";

RL Insect Biochem. Mol. Biol. 31:665-677(2001).
 DR EMBL; AB001822; BAB39501.1;
 SQ SEQUENCE 263 AA; 27501 MW; FFA2BF0943F4930C CRC64;

Query Match 63.1%; Score 41; DB 5; Length 263;

Best Local Similarity 60.0%; Pred. No. 70;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 VFDDGGDTVY 11

DB 52 LFDGGDNNIT 61

RESULT 11

Q9KF21 PRELIMINARY: PRT: 295 AA.

DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein BH0676.
 GN BH0676.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Bacillaceae; Bacillus.
 NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kunara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis."
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL; AP001509; BAB04395.1;
 DR InterPro; IPR005218; Cons_Hypoth147.
 DR InterPro; IPR001206; DAGC.
 DR InterPro; IPR003622; DAG_kin_cat.
 DR Pfam; PF00781; DAGKc_1.
 DR ProDom; PD005043; DAG_kin_cat; 1.
 DR SMART; SM00046; DAGKc; 1.
 DR TIGRFAMs; TIGR00147; TIGR00147; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 295 AA; 32456 MW; C9D93D5DF40B290 CRC64;

Query Match 63.1%; Score 41; DB 16; Length 295;

Best Local Similarity 75.0%; Pred. No. 79;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 GDDGTVYO 12

DB 65 GDDGTIYE 72

RESULT 12

Q9ZBK8 PRELIMINARY: PRT: 380 AA.

DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein SCO6466.
 GN SCO6466 OR SC9C7.02.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;

RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE-97000351; PubMed-8843436;
 RA Redenbach M., Kieser H.M., Denapalte D., Elchner A., Cullum J.,
 RA Kinash H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL; AL035161; CAA22714.1; -
 DR InterPro: IPR004381; Cons_hypoth45.
 DR InterPro: IPR003747; Glycerate_kinase.
 DR Pfam: PF02595; DUF168; 1.
 DR TIGRFAMs: TIGR00045; Cons_hypoth45; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 380 AA; 37816 MW; 7131A7C8735124EB CRC64;
 QY 2 VFDDGGDTV 10
 Db 45 VADGGDGTV 53
 RESULT 13
 Q90WFO PRELIMINARY; PRT; 2658 AA.
 AC Q90WFO;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE CgABP260.
 OS Gallus gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SKELETAL MUSCLE;
 RA Oshima K., Morikawa N., Hashimoto Y., Kuribayashi T., Tachikawa M.,
 RA Terasaki A.G., Ohashi K.;
 RT "Full cDNA sequence of cgABP260.";
 RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AB056475; BAB63944.1; -
 DR InterPro: IPR001589; Actbind_actin.
 DR InterPro: IPR001715; Calponin-like.
 DR InterPro: IPR001298; Filamin.
 DR Pfam: PF00307; CH; 2.
 DR Pfam: PF00630; Filamin; 24.

DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
 DR PROSITE: PS50021; CH; 2.
 DR PROSITE: PS50194; FILAMIN_REPEAT; 20.
 SQ SEQUENCE 2658 AA; 280492 MW; 899436B3257F778 CRC64;
 Query Match
 Best Local Similarity 63.1%; Score 41; DB 13; Length 2658;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SVFDGGDGT 9
 Db 1292 SVSDGGDGT 1300
 RESULT 14
 O86806 PRELIMINARY; PRT; 225 AA.
 AC O86806;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein SC05746.
 DE SC05746 OR SC7C7.01.
 OS Streptomyces coelicolor.
 CC Actinomycetales; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Harris D., Taylor K.;
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE-97000351; PubMed-8843436;
 RA Redenbach M., Kieser H.M., Denapalte D., Elchner A., Cullum J.,
 RA Kinash H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL; AL031031; CAA19847.1; -
 DR InterPro: IPR000653; DegT_Dnri_EryC1.
 DR Pfam: PF01041; DegT_Dnri_EryC1; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 225 AA; 24159 MW; CAEB7D9DDE7912C2 CRC64;
 QY 4 DGGDGTVO 12
 Db 11111111
 Query Match
 Best Local Similarity 61.5%; Score 40; DB 16; Length 225;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 127 DGGDGHITQ 135

RESULT 15

OSVT9D9 PRELIMINARY: PRT: 418 AA.

AC 09VTD9: 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE CG6321 protein (LD15494P).

GN CG6321.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxId=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;

RX MEDLINE-20196006; PubMed-10731132;

RA Adams M.D., Celiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H.C., Blazel R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebo J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Relneert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svrlskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,

RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nunco J., Paclebo J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Xu C., Lewis S.E., Rubin G.M., Celiniker S.,

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AE003546; AAF50113.1; -

DR EMBL: AY061197; AAL28745.1; -

DR FlyBase: FBgn0036117; CG6321.

DR InterPro: IPR004839; AminoTransf1/2.

DR InterPro: IPR00408; Reg_chtr-condens.

DR Pfam: PF00155; aminoLtrao_1_2; 1.

DR PROSITE: PS00626; RC01_2; UNKNOWN_1.

SQ SEQUENCE 418 AA; 47413 MW; 257ACA959ECCABEE CRC64;

Query Match 61.5%; Score 40; DB 5; Length 418;

Best Local Similarity 70.0%; Pred. No. 1.7e+02;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OR 2 VFDGGDGTIV 11

:|||||

DB 13 LFDGGDMVY 22

Search completed: June 16, 2003, 12:09:37

Job time: 6.02459 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:56:18 ; Search time 0.762295 Seconds
(without alignments)
652.918 Million cell updates/sec

Title: US-09-986-682B-2
Perfect score: 65
Sequence: 1 SVFDGSGDTVYQ 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	60.0	297	1 BMRU_BACSU	P39074 bacillus su
2	39	60.0	376	1 ACT1_LYTP1	P53465 lytechinus
3	39	60.0	607	1 GDS1_HUMAN	P52306 homo sapien
4	38	58.5	418	1 LE21_THEMA	O9WYC7 thermotoga
5	38	58.5	440	1 YA33_SCHPO	O09716 schizosacch
6	38	58.5	558	1 GDS1_BOVIN	O04173 bos taurus
7	37.5	57.7	473	1 SABC_BACSU	P05655 bacillus su
8	37	56.9	298	1 Y727_METUA	O58137 methanococo
9	37	56.9	381	1 GRK2_ECOLI	P42100 escherichia
10	37	56.9	382	1 GRK_BACSU	P39713 saccharomyc
11	37	56.9	417	1 YAG1_YEAST	P28183 rhodobacter
12	37	56.9	462	1 SAHH_RHODE	P28183 rhodobacter
13	37	56.9	463	1 SAHH_RHODE	P28183 rhodobacter
14	37	56.9	463	1 SAHH_RHODE	P28183 rhodobacter
15	37	56.9	466	1 SAHH_AGR15	O8U999 agrobacteri
16	37	56.9	466	1 SAHH_BRUME	O8Y649 bruceella me
17	37	56.9	466	1 SAHH_RHODE	O98cm3 rhizobium l
18	37	56.9	466	1 SAHH_RHODE	O98cm3 rhizobium l
19	37	56.9	671	1 CHRA_BACSU	P29072 bacillus su
20	37	56.9	1018	1 CONT_HUMAN	O12660 homo sapien
21	37	56.9	1020	1 CONT_MOUSE	P12360 mus musculu
22	37	56.9	1021	1 CONT_RAT	O63198 rattus norv
23	37	56.9	1877	1 PKC5_MOUSE	O04592 mus musculu
24	37	56.9	1877	1 PKC5_MOUSE	P41413 rattus norv
25	36	55.4	115	1 KY2A_HUMAN	P01614 homo sapien
26	36	55.4	135	1 JANA_DROME	P20348 drosophila
27	36	55.4	308	1 PYRB_METAC	O8T112 metanosarc
28	36	55.4	371	1 GRK_NEIMA	P57098 neisseria m
29	36	55.4	371	1 GRK_NEIMA	P57098 neisseria m
30	36	55.4	381	1 GRK1_ECOLI	P77364 escherichia
31	36	55.4	394	1 CARA_DEIRA	O9Y414 delnoccocus
32	36	55.4	421	1 PGIR_MEDSA	O40312 medicago sa
33	36	55.4	496	1 ARRA_BACST	O9s467 bacillus st

34	36	55.4	605	1 DNAX_AQUPY	O86103 aquifex pyr
35	36	55.4	747	1 TR15_ECOLI	O00192 escherichia
36	36	55.4	1355	1 CA21_RANCA	O42350 rana catesb
37	36	55.4	4303	1 PKD1_HUMAN	P98161 homo sapien
38	35.5	54.6	495	1 YEF1_YEAST	P32622 saccharomyc
39	35	53.8	116	1 CUD4_LOCOM1	P21799 locusta mig
40	35	53.8	232	1 VHED_BPT7	P20313 bacterioph
41	35	53.8	232	1 VHED_BPT7	P03696 bacterioph
42	35	53.8	304	1 Y245_SYNY3	P72702 synechocyst
43	35	53.8	325	1 XY1B_BACOV	P49943 bacterioides
44	35	53.8	378	1 GRK_HAEN	P44507 haemophilus
45	35	53.8	380	1 GRK_BACHD	O92592 bacillus ha

ALIGNMENTS

RESULT 1	ID	BMRU_BACSU	STANDARD:	PRT:	297 AA.
AC	P39074:				
DT	01-FEB-1995 (Rel. 31, Created)				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Protein bmr.				
GN	BMRU.				
OS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_TaxID=1423;				
ON	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168 / JH642;				
RX	MEDLINE=97124195; PubMed=8969508;				
RA	Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,				
RA	Kobayashi Y.;				
RT	"Systematic sequencing of the 283 kb 210 degrees-232 degrees region of				
RT	the Bacillus subtilis genome containing the skin element and many				
RT	sporulation genes.";				
RL	Microbiology 142:3103-3111(1996).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RX	MEDLINE=98044033; PubMed=9384377;				
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,				
RA	Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borchert S.,				
RA	Borisov R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,				
RA	Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,				
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,				
RA	Denioet F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,				
RA	Entlin K.D., Errington J., Fabret C., Ferrati E., Fougere D.,				
RA	Fritz C., Fujita M., Fujita Y., Funo S., Gallizi A., Galleron N.,				
RA	Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,				
RA	Guiseppi G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,				
RA	Hilbert B., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,				
RA	Joris H., Karamata D., Kasahara Y., Kjaer-Blaugard M., Klein C.,				
RA	Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,				
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,				
RA	Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,				
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,				
RA	Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,				
RA	Parro V., Pohl T.M., Portetlelle D., Porwollik S., Prescott A.M.,				
RA	Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,				
RA	Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,				
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,				
RA	Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Solido B.,				

```

RA Sasaki M., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Torouchi M., Tamakoshi A., Tanaka T., Terpsita P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,
RA Vial A., Wandut R., Wedler E., Wedler H., Weitzsaecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumslein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis";
RT Nature 390:249-256(1997).
CC -i- SIMILARITY: TO E.COLI YEGS AND TO STNECHOCYSTIS PCC 6803 SLL0036.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or\_send\_an\_email\_to\_license@sib-sib.ch).
CC -----
DR EMBL, L25604; AAB81538.1; -
DR EMBL, D84432; BAA12602.1; -
DR EMBL, Z99116; CAB14331.1; -
DR Subtilist: BG10302; bmrn.
DR InterPro: IPR005218; Cons_hypoth147.
DR InterPro: IPR001206; DAGKc.
DR InterPro: IPR003622; DAG_kin_cat.
DR Pfam: PF00781; DAGKc; 1.
DR Prodom: PD005043; DAG_kin_cat; 1.
DR SMART: SM00046; DAGKc; 1.
DR TIGRfams: TIGR00147; TIGR00147; 1.
KW Complete proteome.
SQ SEQUENCE 297 AA; 32461 MW; 2EB5B9DC35F90CCD CRC64;
OY 5 GGDGVYQ 12
OY I I I I I I I I
Db 66 GGDGVYQ 73

RESULT 2
ACT1_LYTP1 STANDARD: PRT; 376 AA.
ID ACT1_LYTP1 STANDARD: PRT; 376 AA.
AC P53465.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Actin, cytoskeletal 1 (LPC1).
OS Lytechinus pictus (Painted sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Echinozoidea; Echinacea; Temnopleuroidea; Toxopneustidae;
OC Lytechinus.
OC NCBI_TaxID=7653;
RX MEDLINE=95055791; PubMed=7966365;
RA Fang H., Brandhorst B.P.;
RT "Evolution of actin gene families of sea urchins.";
RL J. Mol. Evol. 39:347-356(1994).
CC -i- FUNCTION: ACTINS ARE HIGHLY CONSERVED PROTEINS THAT ARE INVOLVED
CC IN VARIOUS TYPES OF CELL MOTILITY AND ARE UBIQUITOUSLY EXPRESSED
CC IN ALL EUKARYOTIC CELLS.
CC -i- SUBCELLULAR LOCATION: CYTOSKELETA.
CC -i- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL DIFFERENT SPATIAL
CC TERRITORIES OF THE EMBRYO.
CC -i- SIMILARITY: BELONGS TO THE ACTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U09651; AAK53363.1;
DR HSSP: P02570; 2BTF.
DR InterPro: IPR004001; Actin.
DR InterPro: IPR004000; Actin_like.
DR Pfam: PF000022; actin.1.
DR PRINTS: PR00190; ACTIN.
DR SMART: SM00268; ACTIN.1.
DR PROSITE: PS00406; ACTINS.1; FALSE_NEG.
DR PROSITE: PS00432; ACTINS.2; 1.
DR PROSITE: PS01132; ACTINS_ACT LIKE.1.
KW Structural protein; Multigene family; Acetylation.
FT PROPEP 1 2 REMOVED IN MATURE FORM (BY SIMILARITY).
FT CHAIN 3 376 ACTIN, CYTOSKELETAL 1.
FT MOD_RES 3 3 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 376 AA; 41842 MW; 34F9C575A1F65C90 CRC64;

Query Match 60.0%; Score 39; DB 1; Length 376;
Best Local Similarity 70.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 VFDGDDGV 11
DB 153 VFDGDDGVSY 162

RESULT 3
GDSL_HUMAN STANDARD; PRT; 607 AA.
ID GDSL_HUMAN
AC P52306; Q9NZAB; Q9NYM2;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Rap1 GTPase-GDP dissociation stimulator 1 (Smg p21 stimulatory GDP/GTP
DE exchange protein) (SmG GDS protein) (Exchange factor smgGDS).
GN RAP1GDS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=92195658; Pubmed=1549351;
RA Kikuchi A., Kaibuchi K., Hori Y., Nonaka H., Sakoda T.,
RA Kawamura M., Mizuno T., Takai Y.;
RT "Molecular cloning of the human cDNA for a stimulatory GDP/GTP
RT exchange protein for c-Ki-ras p21 and smg p21.";
RL Oncogene 7:289-293(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Vikis H.G., Stewart S., Guan K.;
RT "SmgGDS as an N17Ras binding protein.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Peripheral blood;
RA Hussey D.J., Albanese N.O., Dobrovic A.;
RT "The major isoform of RAP1GDS1 mRNA in peripheral blood mononuclear
RT cells encodes an isoform of smgGDS with 12 armadillo repeats.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: STIMULATES GDP/GTP EXCHANGE REACTION OF A GROUP OF SMALL
CC GTP-BINDING PROTEINS (G PROTEINS) INCLUDING RAP1A/RAP1B, RHOA,
CC RHOB AND C-81-RAS P21, BY STIMULATING THE DISSOCIATION OF GDP FROM
CC AND THE SUBSEQUENT BINDING OF GTP TO EACH SMALL G PROTEIN.
CC -I- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; are
CC produced by alternative splicing.
CC -I- SIMILARITY: CONTAINS 5 ARM REPEATS.
CC -----
CC THIS SWISS-PROT entry is copyright It is produced through a collaboration

```



```

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RL "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC
CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z50728; CAA90588.1; -
DR KW Hypothetical protein: Transmembrane.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 234 254 POTENTIAL.
FT TRANSMEM 390 410 POTENTIAL.
SO SEQUENCE 440 AA; 49579 MW; CACBB9F3EDDF89AC CRC64;

Query Match 58.5%; Score 38; DB 1; Length 440;
Best Local Similarity 54.5%; Pred. NO. 39;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0.

QY 1 SVPDGDGTVY 11
Db 126 SLFDGGSGPLF 136

RESULT 6
GDS1_BOVIN STANDARD: PRT; 558 AA.
ID GDS1_BOVIN
AC Q04173; Q9TS36;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Rapi GTPase-GDP dissociation stimulator 1 (SMG p21 stimulatory GDP/GTP
DE exchange protein) (SMG GDS protein) (Exchange factor smgGDS).
GN RAP1GDS1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
CX NCBI_TaxID:9913;
OX [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP TISSUE=Brain;
RC MEDLINE=91203910; PubMed=1901951;
RA Kaibuchi K., Mizuno T., Fujioka H., Yamamoto T., Kishi K.,
RA Fukumoto Y., Hori Y., Takai Y.,
RT "Molecular cloning of the cDNA for stimulatory GDP/GTP exchange
RT protein for smg p21s (ras p21-like small GTP-binding proteins) and
RT characterization of stimulatory GDP/GTP exchange protein.";
RL Mol. Cell. Biol. 11:2873-2880(1991).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=92356593; PubMed=1495270;
RA Yamamoto T.;
RT "The stimulatory GDP/GTP exchange protein for ras p21-related small
RT GTP-binding proteins.";
RL Kobe J. Med. Sci. 38:37-56(1992).
RN [3]
RN SEQUENCE OF 350-356.
RP TISSUE=Brain;
RC MEDLINE=90375538; PubMed=2118909;
RA Yamamoto T., Kaibuchi K., Mizuno T., Hiroyoshi M., Shitrateki H.,
RA Takai Y.;
RT "Purification and characterization from bovine brain cytosol of
RT proteins that regulate the GDP/GTP exchange reaction of smg p21s, ras
RT p21-like GTP-binding proteins.";

```

```

CC      -1- SIMILARITY: CONTAINS 3 ARM REPEATS.
CC      -----
CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M63325; AAA21876.1; .
DR      InterPro: IPR000225; Armadillo.
DR      Pfam: PF00514; Armadillo_seg; 5.
DR      SMART; SM00185; ARM; 4.
DR      PROSITE; PS50176; ARM_REPEAT; 2.
KW      GTPase activation; Repeat.
FT      REPEAT 79 118 ARM 1.
FT      REPEAT 121 162 ARM 2.
FT      REPEAT 298 341 ARM 3.
FT      CONFLICT 262 325
FT      FT
FT      CONFLICT 387 450
FT      FT
FT      CONFLICT 387 450
SQ      SEQUENCE 558 AA; 61066 MM; C04C5F435B9B297F CRC64;
QY      2 VFDDGGDTGYVQ 12
Db      260 LFEGGKGWVFQ 270

RESULT 7
SACB_BACSU STANDARD: PRT: 473 AA.
AC P05655; P70984;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE levensucrase precursor (EC 2.4.1.10) (beta-D-fructofuranosyl
DE transferase) (Sucrose 6-fructosyl transferase).
OS BAC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=85295507; PubMed=2993818;
RA Steinmetz M., Le Coq D., Aymerich S., Gonzy-Treboul G., Gay P.;
RT "The DNA sequence of the gene for the secreted Bacillus subtilis
RT enzyme levensucrase and its genetic control sites.";
RL Mol. Gen. Genet. 200:220-228(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Denizot F.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

```



```

DF 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE glycinate kinase 2 (EC 2.7.1.31).
GN GAKR OR B3124.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91154140; PubMed=1705543;
RA Komine Y., Inokuchi H.;
RT "Precise mapping of the rnpB gene encoding the RNA component of RNase
RL P. in Escherichia coli K-12."
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1232-1244(1997).
CC -1- CATALYTIC ACTIVITY: ATP + (R)-glycerate = ADP + 3-phospho-(R)-
CC glycerate.
CC -1- SIMILARITY: BELONGS TO THE GLYCERATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D90212; BAA14239.1; ALT_INIT.
CC EMBL: U18997; AAA57927.1; ALT_INIT.
CC EMBL: AE000394; AAC76158.1; ALT_INIT.
CC PIR: J00614; J00614.
CC DR Ecogen: EG11175; gark.
CC DR InterPro: IPR004381; Cons_hypoth45.
CC DR InterPro: IPR003747; Glycerate_kinase.
CC DR Pfam: PF02595; DUF168.1.
CC DR TIGRfams: TIGR00045; Cons_hypoth45; 1.
CC DR Transfaser: Kinase; Complete proteome.
CC KW SEQUENCE 381 AA; 39104 MW; 0A048E2E5F9FE32 CRC64;
SQ

```

```

Query Match      56.9%; Score 37; DB 1; Length 381;
Best Local Similarity 77.8%; Pred. No. 49;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 2 VFDGDTGV 10
   1 1111111
DB 41 VADGEGTV 49

RESULT 10
GRK_BACSU STANDARD; PRT; 382 AA.
AC P42100.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glycinate kinase (EC 2.7.1.31).
GN GLK OR S14A.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / BGSC1A1;

```

```

RX MEDLINE=96093926; PubMed=7584049;
RA Yoshida K.-I., Seki S., Fujimura M., Miwa Y., Fujita Y.;
RT "Cloning and sequencing of a 36-kb region of the Bacillus subtilis
RL genome between the gnt and iol operons."
RN DNA Res. 2:61-69(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunz F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolyon A., Borcher S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritsch C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koettler P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle R., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serio P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takauchi M., Tamakoshi A., Tanaka T., Terpestra P., Toironi A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarelli A.,
RA Viarri A., Wambert R., Wedler H., Wedler H., Welleneger T.,
RA Winters P., Wipal A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein E.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RL subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 1-13 FROM N.A.
RX MEDLINE=87008613; PubMed=3020045;
RA Fujita Y., Fujita T., Miwa Y., Ninashi J., Aratani Y.;
RT "Organization and transcription of the gluconate operon, gnt, of
RT Bacillus subtilis."
RL J. Biol. Chem. 261:13744-13753(1986).
CC -1- CATALYTIC ACTIVITY: ATP + (R)-glycerate = ADP + 3-phospho-(R)-
CC glycerate.
CC -1- SIMILARITY: BELONGS TO THE GLYCERATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AB005554; BAA21580.1;
CC EMBL: Z99124; CAB16041.1;
CC DR EMBL: X03584; NOT_ANNOTATED_CDS.
CC DR EMBL: X03510; NOT_ANNOTATED_CDS.
CC DR Subtilast: BG1103; glxK.
CC DR InterPro: IPR004381; Cons_hypoth45.
CC DR InterPro: IPR003747; Glycerate_kinase.
CC DR Pfam: PF02595; DUF168.1.
CC DR TIGRfams: TIGR00045; Cons_hypoth45; 1.
CC DR Transfaser: Kinase; Complete proteome.
CC KW SEQUENCE 382 AA; 39398 MW; 9242CB53C0DA5EDB CRC64;
SQ

```

```

Query Match      56.9%; Score 37; DB 1; Length 382;
Best Local Similarity 77.8%; Pred. No. 49;

```

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VFDGDDTV 10
11111111
Db 41 VADGSECTV 49

RESULT 11
YAGL_YEAST
ID YAGL_YEAST STANDARD; PRT: 417 AA.
AC P39713;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical zinc-type alcohol dehydrogenase-like protein in GDH3-CNE1
DE intergenic region.
GN YAL061W OR FUN50.

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95249563; Pubmed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC -1- COFACTOR: ZINC (Potential).
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U12980; AAC04973.1; -
CC SGD: S0000057; YAL061W.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR Pfam: PF00107; adh_zinc.1.
DR PROSITE: PS00059; ADH_ZINC: 1.
KW Hypothetical protein; Oxidoreductase; Zinc.
FT METAL 39 39 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 64 64 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 120 120 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 123 123 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 131 131 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 173 173 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 417 AA; 46098 MW; D921CEDBE61D0151 CRC64;

Query Match 56.9%; Score 37; DB 1; Length 417;
Best Local Similarity 63.6%; Pred. No. 54;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 SVFDGDDTVY 11
11111111
Db 254 SIADGGDGFY 264

RESULT 12
SAHH_ROSDE STANDARD; PRT: 462 AA.
AC Q92NA5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adenosylhomocysteine (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHcyase).
GN AHcy.
OS Roseobacter denitrificans (Erythrobacter sp. (strain OCH 114)).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Roseobacter.
OX NCBI_TaxID=2434;
RN [1]
RP SEQUENCE FROM N.A.
RA Nishimura K., Shimada H., Shimen T., Odayashi T., Masuda T., Ohta H.,
RA Takamiya K.;
RT "Photosynthetic regulatory gene cluster in an aerobic photosynthetic
bacterium, Roseobacter denitrificans.";
RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.
CC -1- COFACTOR: NAD (By similarity).
CC -1- PATHWAY: Activated methyl cycle.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AB020211; BAA34645.1; -
CC HSSP: P10760; 1B3R.
DR InterPro: IPR000043; Ado_hcyase.
DR InterPro: IPR000205; NAD_binding.
DR Pfam: PF00670; AdoHcyase.1.
DR TIGRFAMs: TIGR00936; ahcy.1.
DR PROSITE: PS00738; ADOHCYASE_1; 1.
DR PROSITE: PS00739; ADOHCYASE_2; 1.
KW Hydrolase; NAD: One-carbon metabolism.
FT NP_BIND 246 277 NAD (POTENTIAL).
SQ SEQUENCE 462 AA; 50484 MW; 8D8D3A6F3F896FBE CRC64;

Query Match 56.9%; Score 37; DB 1; Length 462;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 DGGDGVY 11
11111111
Db 128 DGGDGVY 135

RESULT 13
SAHH_RHOCA STANDARD; PRT: 463 AA.
AC P28183;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adenosylhomocysteine (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHcyase).
GN AHcy.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SBI003 / St Louis;
RX MEDLINE=92335291; Pubmed=1631127;
RA Sganga M.W., Aksent R.R., Cantoni G.L., Bauer C.E.;
RT "Mutational and nucleotide sequence analysis of S-adenosyl-L-
homocysteine hydrolase from Rhodobacter capsulatus.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6328-6332(1992).

```

RN [2]
RP SEQUENCE OF 1-13 FROM N.A.
RC STRAIN-SBI003 / St Louis;
RX MEDLINE-94110241; PubMed-8282711;
RA Buggy J.J., Sganga M.W., Bauer C.E.;
RT "Nucleotide sequence and characterization of the Rhodobacter
RT capsulatus hvrB gene: HvrB is an activator of S-adenosyl-L-
RT homocysteine hydrolase expression and is a member of the LysR
RT family.";
RL J. Bacteriol. 176:61-69(1994).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.
CC -1- COFACTOR: NAD.
CC -1- PATHWAY: Activated methyl cycle.
CC -1- SUBUNIT: Homotetramer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE ADENOSYLMOMOCYSTEINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M80630; AAA26094.1; -.
CC EMBL: L23836; AAA53540.1; -.
CC HSSP: P10760; 1B3R.
CC InterPro: IPR000043; Ado_hcyase.
CC Pfam: PF00670; AdoHcyase; 1.
CC TIGRfams: TIGR00936; ancy; 1.
CC PROSITE: PS00738; ADOHCYASE_1; 1.
CC PROSITE: PS00739; ADOHCYASE_2; 1.
CC Hydrolyase: NAD: One-carbon metabolism.
CC NP_BIND: 247 278 NAD (POTENTIAL).
CC SEQUENCE 463 AA; 50580 MW; E4EB19B320420B53 CRC64;
SQ
Query Match 56.9%; Score 37; DB 1; Length 463;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 4 DGGDCTV 11
Db 128 DGGDAILY 135
RESULT 14
SAHH_RHOSH STANDARD: PRT; 463 AA.
ID SAHH_RHOSH
AC O50562;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adenosylhomocysteine (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHcyase).
GN AHcy.
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
CC Rhodobacter.
CC NCB1_TaxID=1063;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE-97354111; PubMed-9210332;
RX Mitoquchi H., Masuda T., Nishimura K., Shimada H., Ohta H., Shioi Y.,
RA Takamiya K.;
RT "Nucleotide sequence and transcriptional analysis of the flanking
RT region of the gene (spb) for the trans-acting factor that controls
RT light-mediated expression of the puf operon in Rhodobacter
RT sphaeroides.";
RL Plant Cell Physiol. 38:558-567(1997).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.

```

```

CC -1- COFACTOR: NAD (By similarity).
CC -1- PATHWAY: Activated methyl cycle.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE ADENOSYLMOMOCYSTEINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U76671; AAB88245.1; -.
CC HSSP: P10760; 1B3R.
CC InterPro: IPR000043; Ado_hcyase.
CC Pfam: PF00670; AdoHcyase; 1.
CC TIGRfams: TIGR00936; ancy; 1.
CC PROSITE: PS00738; ADOHCYASE_1; 1.
CC PROSITE: PS00739; ADOHCYASE_2; 1.
CC Hydrolyase: NAD: One-carbon metabolism.
CC NP_BIND: 247 278 NAD (POTENTIAL).
CC SEQUENCE 463 AA; 50617 MW; ACE8EB9D0D0CCBBA CRC64;
SQ
Query Match 56.9%; Score 37; DB 1; Length 463;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 4 DGGDCTV 11
Db 128 DGGDAILY 135
RESULT 15
SAHH_AGR5 STANDARD: PRT; 466 AA.
ID SAHH_AGR5
AC Q80099;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adenosylhomocysteine (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHcyase).
GN AHcy OR ARU0029 OR AGR_C_46.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Rhizobiaceae; Rhizobium.
CC NCB1_TaxID=176299;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE-21608550; PubMed-11743193;
RX Wood D.W., Setubal J.C., Kaul R., Almeida N.F., Woo L.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Wester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RN SEQUENCE FROM N.A.
RP MEDLINE-21608551; PubMed-11743194;
RX Goodner B., Hinkle G., Gattung S., Askenazi M., Halling C., Mullin L.,
RA Quicillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Irtchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughly D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursun J., Lomo C., Sear C., Strub G.,
RA Cleto C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent

```

```
RT Agrobacterium tumefaciens C58.
RL Science 294:2323-2328(2001).
CC -I- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.
CC -I- COFACTOR: NAD (By similarity).
CC -I- PATHWAY: Activated methyl cycle.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008977; AAA1060.1; -
DR EMBL; AE007946; AAK8583.1; -
DR PROSITE; PS00738; ADOHCYASE_1; 1.
DR PROSITE; PS00739; ADOHCYASE_2; 1.
KW Hydrolase; NAD; One-carbon metabolism; Complete proteome.
FT NP BIND 250 281 NAD (PARENTIAL).
SQ SEQUENCE 466 AA; 51159 MW; 9A124B3304419F26 CRC64;

Query Match 56.9%; Score 37; DB 1; Length 466;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 DGGDGTIV 11
   1111 1:1
Db 132 DGGDATMY 139
```

Search completed: June 16, 2003, 12:07:25
Job time : 1.7623 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 16, 2003, 12:03:38 ; Search time 1.47541 Seconds
(without alignments)
781.893 Million cell updates/sec

Title: US-09-986-682B-2
Perfect score: 65
Sequence: 1 SVFDGNGTIVQ 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	59	90.8	489	2 F97118	levansucrase [impo
2	43	66.2	330	2 D87393	hypothetical prote
3	42	64.6	767	2 T30018	hypothetical prote
4	41	63.1	295	2 D83734	hypothetical prote
5	41	63.1	380	2 T35953	conserved hypothe
6	40	61.5	225	2 T35679	hypothetical prote
7	40	61.5	469	2 T35670	hypothetical prote
8	39	60.0	108	2 B86321	hypothetical prote
9	39	60.0	297	2 F69595	multidrug resistan
10	39	60.0	523	2 T11317	cytochrome-c oxida
11	39	60.0	518	2 I37456	smg GDS - human
12	39	60.0	655	2 JEO358	raib binding prote
13	38.5	59.2	647	2 A84265	hypothetical prote
14	38.5	59.2	815	2 T40524	hypothetical prote
15	38	58.5	146	2 C86187	hypothetical prote
16	38	58.5	382	2 T49762	hypothetical prote
17	38	58.5	388	2 J70308	cellulase (EC 3.2.
18	38	58.5	398	2 H70393	hemolysin - Aquife
19	38	58.5	418	2 C72394	hypothetical prote
20	38	58.5	440	2 S58302	hypothetical prote
21	38	58.5	558	2 A38795	stimulatory GDP/GT
22	38	58.5	1192	2 T48499	receptor-like prot
23	37.5	57.7	473	2 A25040	levansucrase (EC 2
24	37	56.9	82	2 S65779	gamma-thionin prec
25	37	56.9	249	2 G84224	hypothetical prote
26	37	56.9	298	2 G64390	coenzyme F420 hydr
27	37	56.9	311	2 F83871	multidrug resistan
28	37	56.9	342	2 G72567	hypothetical prote
29	37	56.9	377	2 E82402	conserved hypothet

30	37	56.9	379	2 C75167	hypothetical prote
31	37	56.9	381	2 AD0897	conserved hypothet
32	37	56.9	382	2 A70071	conserved hypothet
33	37	56.9	387	2 B85974	hypothetical prote
34	37	56.9	408	2 J00614	hypothetical prote
35	37	56.9	408	2 B91129	Yhad protein - Esc
36	37	56.9	417	2 S51961	hypothetical prote
37	37	56.9	462	1 A46035	FUN50 protein - ye
38	37	56.9	466	2 D97362	adenosylhomocystei
39	37	56.9	466	2 AP2580	adenosylhomocystei
40	37	56.9	481	2 AG3505	S-adenosylhomocyst
41	37	56.9	620	2 T30765	adenosylhomocystei
42	37	56.9	629	2 H86383	hypothetical prote
43	37	56.9	671	1 ORSCN	probable wall-asso
44	37	56.9	747	2 A71440	two-component sens
45	37	56.9	801	2 G82302	hypothetical prote
					probable cellobios

ALIGNMENTS

```
RESULT 1
F97118
levansucrase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: F97118
R.Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
., Daly, M.D.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97118
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-489 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79737.1; PID:q15024742; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1772

Query Match
Best Local Similarity 90.8%; Score 59; DB 2; Length 489;
Matches 10; Conservative 83.3%; Pred. No. 0.031;
Mismatches 2; Indels 0; Gaps 0;

QY 1 SVFDGNGTIVQ 12
| | | | | | | | | | | |
Db 229 SVFDGNGTIVQ 240

RESULT 2
D87393
hypothetical protein CC1160 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: D87393
R.Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Dirkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Fernalda, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: D87393
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-330 <STO>
A:Cross-references: GB:AE005673; NID:q13422480; PIDN:AAK23144.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1160

Query Match
Best Local Similarity 66.2%; Score 43; DB 2; Length 330;
Matches 8; Conservative 88.9%; Pred. No. 10;
Mismatches 0; Indels 1; Gaps 0;
```

QY 2 VFDDGDTGV 10
| | | | |
Db. 80 VIDGGDTGV 88

RESULT 3

T30018
hypothetical protein F38B9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T30018
R:Wu, X.; Gattung, S.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F38B9.
A:Reference number: 220722
A:Accession: T30018
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-767 <WUX>

A:Cross-references: EMBL:U46668; PIDN:AAA93348.1; CESP:F38B9.2

C:Genetics:

A:Gene: CESP:F38B9.2

A:Introns: 13/1; 37/3; 62/1; 89/1; 127/1; 164/2; 182/1; 235/3; 269/3; 321/2; 360/1; 455/

Query Match 64.6%; Score 42; DB 2; Length 767;

Best Local Similarity 77.8%; Pred. No. 36;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVFDGDTGV 9
| | | | |

Db 592 SYDDGDTGV 600

RESULT 4

D83734

hypothetical protein BH0676 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: D83734

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirz
Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: D83734

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-295 <STO>

A:Cross-references: GB:AP001509; GB:BA000004; NID:q10173176; PIDN:BAH04395.1; GSPDB:GNOC

C:Genetics:

A:Experimental source: strain C-125

A:Gene: BH0676

QY 5 GDDGTGVQ 12
| | | | |

Db 65 GDDGTGV 72

RESULT 5

T35953

conserved hypothetical protein SC9C7.02 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000

C:Accession: T35953

R:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1999

A:Reference number: 221551

A:Accession: T35953

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-380 <SEE>
A:Cross-references: EMBL:AL035161; PIDN:CAA22714.1; GSPDB:GN00070; SCOEDB:SC9C7.02
A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC9C7.02

C:Superfamily: yhad protein

Query Match 63.1%; Score 41; DB 2; Length 380;

Best Local Similarity 88.9%; Pred. No. 26;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VFDDGDTGV 10
| | | | |

Db 45 VADGGDTGV 53

RESULT 6

T35679
hypothetical protein SC7C7.01 SC7C7.01 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T35679

R:Harris, D.; Taylor, K.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1998

A:Reference number: 221587

A:Accession: T35679

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-225 <HAR>

A:Cross-references: EMBL:AL031031; PIDN:CAA19847.1; GSPDB:GN00070; SCOEDB:SC7C7.01

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC7C7.01

Query Match 61.5%; Score 40; DB 2; Length 225;

Best Local Similarity 77.8%; Pred. No. 22;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 DGGDTGVQ 12
| | | | |

Db 127 DGGDTGVQ 135

RESULT 7

T35670

hypothetical protein SC7B7.02 SC7B7.02 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T35670

R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, November 1997

A:Reference number: 221586

A:Accession: T35670

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-469 <MUR>

A:Cross-references: EMBL:AL009199; PIDN:CAA15782.1; GSPDB:GN00070; SCOEDB:SC7B7.02

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC7B7.02

Query Match 61.5%; Score 40; DB 2; Length 469;

Best Local Similarity 63.6%; Pred. No. 48;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SVFDGDTGV 11
| | | | |

Db 65 AFDGGDTGV 75

RESULT 8

B86321

hypothetical protein F6A14.16 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: B86321
R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Lin, Y.; Lin, X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: B86321
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-108 <STO>
A:Cross-references: GB:AE005172; NID:96730711; PIDN:AAE27106.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 60.0%; Score 39; DB 2; Length 108;
Best Local Similarity 87.5%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 DGGDGTVO 11
DB 59 DGGDGTVO 66

RESULT 9
F69595

multidrug resistance protein cotranscribed with bmr bmr - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: F69595
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter,
C.; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferratti, E.
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Galler
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schaefer, R.; Scifone, F.; Sekiguchi, J.; Sekowska, A.; Serot
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstira, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zunshtein, E.; Yoshikawa, H.; Danculin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:96044033; PMID:9384377
A:Accession: F69595
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-297 <KUN>
A:Cross-references: GB:299116; GB:AL009126; NID:92634723; PIDN:CA814331.1; PID:ell185668;
A:Experimental source: strain 168
C:Genetics:
A:Gene: bmr

Query Match 60.0%; Score 39; DB 2; Length 297;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 5 GGDGTVO 12
DB 66 GGDGTVO 73

RESULT 10

T11317

cytochrome-c oxidase (EC 1.9.3.1) chain I - Pedinomonas minor mitochondrion
C:Species: mitochondrion Pedinomonas minor
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
C:Accession: T11317
R:Turner, M.; Lemieux, C.; Burger, G.; Lang, B.F.; Otis, C.; Plante, I.; Gray, M.W.
submitted to the EMBL Data Library, December 1998
A:Description: The complete mitochondrial DNA sequences of Nephrolepis olivacea and
A:Reference number: Z17261

A:Accession: T11317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-523 <TUR>
A:Cross-references: EMBL:AF116775; NID:94378766; PID:94378771; PIDN:AA019669.1
C:Genetics:

A:Gene: cox1
A:Genome: mitochondrion
A:Genetic code: SGC3

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated com
F:12-459/Domain: cytochrome-c oxidase chain I homology <CO1>

F:64,380/Binding site: heme a iron (His) (axial ligands) #status predicted
F:243,292/Binding site: copper (His) #status predicted
F:243,247/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:247/Binding site: oxygen (Tyr) #status predicted
F:370/Binding site: magnesium (His) (shared with chain II) #status predicted
F:378/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 60.0%; Score 39; DB 2; Length 523;
Best Local Similarity 58.3%; Pred. No. 78;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 SVFDGGDGTVO 12
DB 224 SVFDGGDGTVO 235

RESULT 11
137456

smg GDS - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change 05-Nov-1999
C:Accession: I37456; S22144
R:Kikuchi, A.; Kikuchi, K.; Horii, Y.; Nonaka, H.; Sakoda, T.; Kawamura, M.; Mizuno,
Oncogene 7, 289-293, 1992
A:Title: Molecular cloning of the human cDNA for a stimulatory GDP/GTP exchange prote
A:Reference number: I37456; MUID:92195658; PMID:1549351
A:Accession: I37456
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-558 <RES>
A:Cross-references: EMBL:X63465; NID:932079; PIDN:CAA45067.1; PID:932080
C:Genetics:

A:Gene: hgdS
Query Match 60.0%; Score 39; DB 2; Length 558;
Best Local Similarity 54.5%; Pred. No. 84;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 VFDGGDGTVO 12
DB 260 VFDGGDGTVO 270

RESULT 12
JE0358

ralB binding protein - clawed frog
C:Species: Xenopus sp. (clawed frog)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 07-May-1999
C:Accession: JE0358
R:Iouzalet, N.; Camonis, J.; Moreau, J.
Biochem. Biophys. Res. Commun. 250, 359-363, 1998
A:Title: Identification and characterization in Xenopus of XsmgGDS, a RalB binding pr

A:Reference number: JE0358; MUID:98440802; PMID:9753634
 A:Accession: JE0358
 A:Molecule type: mRNA
 A:Residues: 1-615 <IOU>
 A:Cross-references: GB:AJ005870

Query Match 60.0%; Score 39; DB 2; Length 615;
 Best Local Similarity 54.5%; Pred. No. 93;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 VFDDGDTFTVQ 12
 :|||1:|:|
 Db 317 LFEKGKGVFQ 327

RESULT 13
 A84265
 hypothetical protein Vng1087c [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: A84265
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leitner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo, Jung, K.H.; Alam, M.; Freitas, T.; 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li, A.; Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; MUID:20504483; PMID:11016950
 A:Accession: A84265
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-647 <STO>
 A:Cross-references: GB:AE004437; NID:q10580635; PIDN:AA619485.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: VNG1087C

Query Match 59.2%; Score 38.5; DB 2; Length 647;
 Best Local Similarity 50.0%; Pred. No. 1.2e+02;
 Matches 9; Conservative 1; Mismatches 1; Indels 7; Gaps 1;

OY 2 VFDDGDTFTVQ 12
 :|||1:|:|
 Db 53 VFDDGMDRVADADGTYE 70

RESULT 14
 T40524
 hypothetical protein SPBC530.08 - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
 C:Accession: T40524
 R:Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M. submitted to the EMBL Data Library, May 1998
 A:Reference number: Z21934
 A:Accession: T40524
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-815 <LYN>
 A:Cross-references: EMBL:AL023634; PIDN:CAA19174.1; GSPDB:GN00067; SPDB:SPBC530.08
 C:Genetics:
 A:Gene: SPDB:SPBC530.08
 A:Map position: 2
 C:Superfamily: GAL4 zinc binuclear cluster homology
 F:26-62/Domain: GAL4 zinc binuclear cluster homology <GL4>

Query Match 59.2%; Score 38.5; DB 2; Length 815;
 Best Local Similarity 69.2%; Pred. No. 1.5e+02;
 Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

OY 1 SVFDGDTFTV-Q 12
 :|||1:|:|
 Db 680 SVFDGNTGTTIPYQ 692

RESULT 15
 C86187
 YUP8H12.12 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: C86187
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, Jensen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Matli, R.; Marzia, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: C86187
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-146 <STO>
 A:Cross-references: GB:AE005172; NID:q2388570; PIDN:AA871451.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 58.5%; Score 38; DB 2; Length 146;
 Best Local Similarity 87.5%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SVFDGDTG 8
 :|||1:|:|
 Db 61 SVFDGDTG 68

Search completed: June 16, 2003, 12:10:42
 Job time : 3.47541 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:07:33 ; Search time 1.84426 seconds
(without alignments)
695.665 Million cells updates/sec

Title: US-09-986-682B-2
Perfect score: 65
Sequence: 1 SVFDGSDGVYQ 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PublishedApplications_AA.*
1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppaa/PCT07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppaa/PCT07_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	100.0	12	10	US-09-986-682B-2
2	65	100.0	455	10	US-09-986-682B-3
3	38	58.5	295	9	US-10-223-085-370
4	38	58.5	295	9	US-10-223-084-370
5	38	58.5	295	9	US-10-223-088-370
6	38	58.5	295	9	US-10-223-090-370
7	37	56.9	408	10	US-09-912-020-235
8	37	56.9	811	10	US-09-815-242-11875
9	37	56.9	1548	9	US-10-180-903-2
10	37	56.9	2076	10	US-09-815-242-5815
11	37	56.9	2186	10	US-09-815-242-11913
12	36	55.4	37	9	US-09-981-876-230
13	36	55.4	37	9	US-09-148-545-230
14	36	55.4	113	10	US-09-840-459-68
15	36	55.4	375	10	US-09-815-242-13314
16	35	53.8	4303	9	US-09-904-968A-2
17	35	53.8	140	9	US-10-101-464A-688
18	35	53.8	361	9	US-09-738-626-3735
19	35	53.8	378	9	US-10-260-877-48

20	35	53.8	382	10	US-09-815-242-10502	Sequence 10502, A
21	35	53.8	426	9	US-09-815-242-10569	Sequence 10569, A
22	35	53.8	499	9	US-09-738-626-4342	Sequence 4342, Ap
23	35	53.8	503	10	US-09-990-337-3	Sequence 3, Appl
24	35	53.8	589	12	US-10-001-851-26	Sequence 26, Appl
25	35	53.8	618	12	US-10-005-647-2	Sequence 2, Appl
26	35	53.8	627	9	US-10-050-704-125	Sequence 125, App
27	35	53.8	953	9	US-09-884-696-3	Sequence 3, Appl
28	35	53.8	956	9	US-10-101-664A-895	Sequence 4, Appl
29	35	53.8	998	9	US-09-947-063-13	Sequence 895, App
30	35	53.8	1028	9	US-09-947-063-13	Sequence 13, Appl
31	35	53.8	1028	9	US-09-947-063-14	Sequence 14, Appl
32	35	53.8	1311	9	US-10-103-377C-6	Sequence 6, Appl
33	35	53.8	4349	9	US-10-160-758-15	Sequence 15, Appl
34	34	52.3	79	10	US-09-939-980-403	Sequence 403, App
35	34	52.3	81	10	US-09-815-242-12719	Sequence 12719, A
36	34	52.3	133	12	US-10-006-773-9	Sequence 9, Appl
37	34	52.3	255	9	US-09-898-837A-36	Sequence 36, Appl
38	34	52.3	260	10	US-09-731-231A-5	Sequence 5, Appl
39	34	52.3	267	10	US-09-815-242-5773	Sequence 5773, Ap
40	34	52.3	283	10	US-09-952-689-5	Sequence 5, Appl
41	34	52.3	303	10	US-09-815-242-5393	Sequence 5393, Ap
42	34	52.3	306	10	US-09-815-242-12300	Sequence 12300, A
43	34	52.3	311	9	US-09-988-626-230	Sequence 230, App
44	34	52.3	311	9	US-09-988-687-230	Sequence 230, App
45	34	52.3	311	9	US-10-190-279-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-09-986-682B-2
Sequence 2, Appl
Patent No. US20020115182A1
GENERAL INFORMATION:
APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU
KINKYUJO
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/986,682B
FILING DATE: 06-NO. US20020115182A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
TOPOLOGY: linear
FRAGMENT TYPE: internal fragment

SEQUENCE DESCRIPTION: SEQ ID NO: 2
US-09-986-682B-2

Query Match 100.0%; Score 65; DB 10; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVFDGDDGTVYQ 12
Db 1 SVFDGDDGTVYQ 12

RESULT 2
US-09-986-682B-3
Sequence 3, Application US/09986682B
Patent No. US20020115182A1
GENERAL INFORMATION:
APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU
KENRYUO
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOPURANOSIDASE ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/986, 682B
FILING DATE: 06-NO. US20020115182A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-986-682B-3

Query Match 100.0%; Score 65; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVFDGDDGTVYQ 12
Db 201 SVFDGDDGTVYQ 212

RESULT 3
US-10-223-085-370
Sequence 370, Application US/10223085
Publication No. US20030100497A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Ferrara, Napoleone

APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Masters, Scot A.
APPLICANT: Pan, James
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Ye, Weilian
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: TREATMENT OF DISORDERS INVOLVING ANGIOGENESIS
FILE REFERENCE: P3235P1C10
CURRENT APPLICATION NUMBER: US/10/223,085
CURRENT FILING DATE: 2002-08-16
PRIOR APPLICATION NUMBER: US 10/081,056
PRIOR FILING DATE: 2002-02-20
PRIOR APPLICATION NUMBER: US 60/213,637
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/219,556
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: US 60/220,624
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/220,664
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: US 60/222,695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: US 09/643,657
PRIOR FILING DATE: 2000-08-17
PRIOR APPLICATION NUMBER: PCT/US00/23522
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US00/23328
PRIOR FILING DATE: 2000-08-24
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 383
SEQ ID NO 370
LENGTH: 295.
TYPE: PRT
ORGANISM: Homo sapiens
US-10-223-085-370

Query Match 58.5%; Score 38; DB 9; Length 295;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 FDGDDGTV 10
Db 58 FDGDDATI 65

RESULT 4
US-10-223-084-370
Sequence 370, Application US/10223084
Publication No. US20030105011A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Ferrara, Napoleone
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Masters, Scot A.
APPLICANT: Pan, James
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.

```
APPLICANT: Williams, P. Mickey
APPLICANT: Ye, Weilan
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: TREATMENT OF DISORDERS INVOLVING ANGIOGENESIS
FILE REFERENCE: P3235PIC5
CURRENT APPLICATION NUMBER: US/10/223,084
CURRENT FILING DATE: 2002-08-16
PRIOR APPLICATION NUMBER: US 10/081,056
PRIOR FILING DATE: 2002-02-20
PRIOR APPLICATION NUMBER: US 60/213,637
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/219,556
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: US 60/220,624
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/220,664
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: US 60/222,695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: US 09/643,657
PRIOR FILING DATE: 2000-08-17
PRIOR APPLICATION NUMBER: PCT/US00/23522
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US00/23328
PRIOR FILING DATE: 2000-08-24
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 383
SEQ ID NO 370
LENGTH: 295
TYPE: PRT
ORGANISM: Homo sapiens
US-10-223-084-370
```

```
Query Match      58.5%; Score 38; DB 9; Length 295;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      3 FDGSDGTV 10
        ||||| 1:
Db      58 FDGSDATI 65
```

RESULT 5

```
US-10-223-088-370
Sequence 370, Application US/10223088
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Ferrara, Napoleone
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Marsters, Scott A.
APPLICANT: Pan, James
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Williams, P. Mickey
APPLICANT: Ye, Weilan
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: TREATMENT OF DISORDERS INVOLVING ANGIOGENESIS
FILE REFERENCE: P3235PIC6
CURRENT APPLICATION NUMBER: US/10/223,088
CURRENT FILING DATE: 2002-08-16
PRIOR APPLICATION NUMBER: US 10/081,056
PRIOR FILING DATE: 2002-02-20
PRIOR APPLICATION NUMBER: US 60/213,637
PRIOR FILING DATE: 2000-06-23
```

```
PRIOR APPLICATION NUMBER: US 60/219,556
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: US 60/220,624
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/220,664
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: US 60/222,695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: US 09/643,657
PRIOR FILING DATE: 2000-08-17
PRIOR APPLICATION NUMBER: PCT/US00/23522
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US00/23328
PRIOR FILING DATE: 2000-08-24
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 383
SEQ ID NO 370
LENGTH: 295
TYPE: PRT
ORGANISM: Homo sapiens
US-10-223-088-370
```

```
Query Match      58.5%; Score 38; DB 9; Length 295;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      3 FDGSDGTV 10
        ||||| 1:
Db      58 FDGSDATI 65
```

RESULT 6

```
US-10-223-090-370
Sequence 370, Application US/10223090
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Ferrara, Napoleone
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Marsters, Scott A.
APPLICANT: Pan, James
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Williams, P. Mickey
APPLICANT: Ye, Weilan
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: TREATMENT OF DISORDERS INVOLVING ANGIOGENESIS
FILE REFERENCE: P3235PIC2
CURRENT APPLICATION NUMBER: US/10/223,090
CURRENT FILING DATE: 2002-08-16
PRIOR APPLICATION NUMBER: US 10/081,056
PRIOR FILING DATE: 2002-02-20
PRIOR APPLICATION NUMBER: US 60/213,637
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/219,556
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: US 60/220,624
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: US 60/222,695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: US 09/643,657
```

;; PRIOR FILING DATE: 2000-08-17
;; PRIOR APPLICATION NUMBER: PCT/US00/23522
;; PRIOR FILING DATE: 2000-08-23
;; PRIOR APPLICATION NUMBER: PCT/US00/23328
;; PRIOR FILING DATE: 2000-08-24
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 383
;; SEQ ID NO 370
;; LENGTH: 295
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-223-090-370

Query Match 58.5%; Score 38; DB 9; Length 295;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 FDGDDGTIV 10
Db 58 FDGDDATI 65

RESULT 7
US-09-912-020-255
;; Sequence 255, Application US/09912020
;; Patent No. US20020045592A1
;; GENERAL INFORMATION:
;; APPLICANT: Ohlsen, Karl L.
;; APPLICANT: Trawick, John
;; APPLICANT: Forsyth, R. Allyn
;; APPLICANT: Froelich, Jamie M.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.
;; APPLICANT: Xu, H. Howard
;; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;; FILE REFERENCE: ELITRA.001DV1
;; CURRENT APPLICATION NUMBER: US/09/912,020
;; CURRENT FILING DATE: 2001-07-23
;; PRIOR APPLICATION NUMBER: 09/492,709
;; PRIOR FILING DATE: 2000-01-27
;; PRIOR APPLICATION NUMBER: 60/117,405
;; PRIOR FILING DATE: 1999-01-27
;; NUMBER OF SEQ ID NOS: 485
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 255
;; LENGTH: 408
;; TYPE: PRT
;; ORGANISM: E. Coli
US-09-912-020-255

Query Match 56.9%; Score 37; DB 10; Length 408;
Best Local Similarity 77.8%; Pred. No. 2.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 VPDGDDGTIV 10
Db 68 -VADGGEGTV 76

RESULT 8
US-09-815-242-11875
;; Sequence 11875, Application US/09815242
;; Patent No. US20020061569A1
;; GENERAL INFORMATION:
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Karl L.
;; APPLICANT: Zyskind, Judith W.
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John D.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.

;; APPLICANT: Xu, H. Howard
;; TITLE OF INVENTION: Identification of Essential Genes in
;; FILE REFERENCE: ELITRA.011A
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; CURRENT FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 1410
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 11875
;; LENGTH: 811
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11875

Query Match 56.9%; Score 37; DB 10; Length 811;
Best Local Similarity 85.7%; Pred. No. 5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 FDGDDGT 9
Db 734 FDGDDGS 740

RESULT 9
US-10-180-903-2
;; Sequence 2, Application US/10180903
;; Publication No. US2003009382A1
;; GENERAL INFORMATION:
;; APPLICANT: Allen, Keith D.
;; TITLE OF INVENTION: SPC6 SERINE PROTEASE GENE DISRUPTIONS,
;; FILE REFERENCE: R-720
;; CURRENT APPLICATION NUMBER: US/10/180,903
;; CURRENT FILING DATE: 2002-06-25
;; PRIOR APPLICATION NUMBER: US 60/300,978
;; PRIOR FILING DATE: 2001-06-26
;; PRIOR APPLICATION NUMBER: US 60/324,820
;; PRIOR FILING DATE: 2001-09-24
;; NUMBER OF SEQ ID NOS: 4
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 2
;; LENGTH: 1548
;; TYPE: PRT
;; ORGANISM: Mus Musculus
US-10-180-903-2

Query Match 56.9%; Score 37; DB 9; Length 1548;
Best Local Similarity 63.6%; Pred. No. 9.9e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPDGDDGTIVQ 12
Db 1513 VYMGDDGTIVR 1523

RESULT 10
US-09-815-242-5815
;; Sequence 5815, Application US/09815242
;; Patent No. US20020061569A1

```

; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5815
; LENGTH: 2076
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5815

Query Match          56.9%; Score 37; DB 10; Length 2076;
Best Local Similarity 58.3%; Pred. No. 1.3e+03;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      1  SVFDGGDGTVYQ 12
       1:1111111111
Db      291  SMFDGDKNKYQ 302

RESULT 11
US-09-815-242-12913
; Sequence 12913, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5815
; LENGTH: 2076
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12913
```

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 12913
; LENGTH: 2186
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12913

Query Match          56.9%; Score 37; DB 10; Length 2186;
Best Local Similarity 58.3%; Pred. No. 1.4e+03;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      1  SVFDGGDGTVYQ 12
       1:1111111111
Db      291  SMFDGDKNKYQ 302

RESULT 12
US-09-981-876-230
; Sequence 230, Application US/09981876
; Patent No. US20020164669A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: PZ001P1
; CURRENT APPLICATION NUMBER: US/09/981,876
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: 09/148,545
; PRIOR FILING DATE: 1998-09-04
; PRIOR APPLICATION NUMBER: 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,333
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/038,621
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,161
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,626
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,334
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,336
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,163
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/047,615
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,600
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,597
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,502
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,633
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,583
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,617
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,618
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,503
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,592
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,581
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,584
; PRIOR FILING DATE: 1997-05-23
```


Query Match: 55.48; Score 36; DB 9; Length 37;
Best Local Similarity 75.08; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 DGGDCTVY 11
|||||
Db 25 DGGDCTSF 32

RESULT 13
US-09-148-545-230
Sequence 230, Application US/09148545
Publication No. US20030027132A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 70 Human Secreted Proteins
FILE REFERENCE: P2001P1
CURRENT APPLICATION NUMBER: US/09/148,545
EARLIER FILING DATE: 1998-09-04
EARLIER APPLICATION NUMBER: PCT/US98/04482
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,161
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598

EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056,886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,911
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,636
EARLIER FILING DATE: 1997-08-22

EARLIER APPLICATION NUMBER: 60/056,874
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,910
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,864
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,631
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,845
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,892
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047,595
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/057,761
EARLIER FILING DATE: 05-Sep-1997
EARLIER APPLICATION NUMBER: 60/047,599
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,588
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,585
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,594
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,589
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
NUMBER OF SEQ ID NOS: 280
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 230
LENGTH: 37

Query Match 55.4%; Score 36; DB 9; Length 37;
Best Local Similarity 75.0%; Pred. No. 28;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 4 DGGDGTGY 11
|||||
Db 25 DGGDGTSF 32

RESULT 14
US-09-840-459-68
Sequence 68, Application US/09840459
Patent No. US20020150576A1
GENERAL INFORMATION:
APPLICANT: Lakosa, Gregory J.
APPLICANT: Horvath, Christopher
APPLICANT: Newman, Walter
APPLICANT: Jones, S. Tarran
APPLICANT: O'Keefe, Theresa
TITLE OF INVENTION: HUMANIZED ANTI-CCR2 ANTIBODIES AND
FILE REFERENCE: 1855,1052-012
CURRENT APPLICATION NUMBER: US/09/840,459
CURRENT FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: PCT/US01/03537
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 09/497,625
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: 09/359,193
PRIOR FILING DATE: 1999-07-22
PRIOR APPLICATION NUMBER: 09/121,781
PRIOR FILING DATE: 1998-07-23
NUMBER OF SEQ ID NOS: 107
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 68
LENGTH: 113
TYPE: PRT
ORGANISM: Homo sapiens
US-09-840-459-68

Query Match 55.4%; Score 36; DB 10; Length 113;
Best Local Similarity 54.5%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 SVFDGDDGTGY 11
|:| ||| |
Db 28 SLDSGDCNTY 38

RESULT 15
US-09-815-242-13314
Sequence 13314, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 13314
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13314

Query Match 55.4%; Score 36; DB 10; Length 375;
Best Local Similarity 77.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 VPDGGDGTV 10
| | | | |
Db 45 VPDGGEGTV 53

Search completed: June 16, 2003, 12:22:36
Job time : 2.84426 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:04:38 ; Search time 1.08197 Seconds
(without alignments)
326.327 Million cell updates/sec

Title: US-09-986-682B-2
Perfect score: 65
Sequence: 1 SVFDGSDGTIVYQ 12

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PCUS-COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	100.0	12	2	US-08-870-827-2 Sequence 2, Appli
2	65	100.0	12	4	US-09-317-179-2 Sequence 2, Appli
3	65	100.0	455	2	US-08-870-827-3 Sequence 3, Appli
4	65	100.0	455	4	US-09-317-179-3 Sequence 3, Appli
5	37	56.9	190	2	US-08-560-398-8 Sequence 8, Appli
6	37	56.9	1018	1	US-08-408-093-6 Sequence 6, Appli
7	37	56.9	1018	1	US-08-408-420A-6 Sequence 6, Appli
8	37	56.9	1018	1	US-08-714-901-6 Sequence 6, Appli
9	37	56.9	1018	1	US-08-452-052-2 Sequence 2, Appli
10	37	56.9	1018	3	US-08-040-741-6 Sequence 6, Appli
11	36	55.4	460	1	US-08-289-709-1 Sequence 1, Appli
12	36	55.4	460	1	US-08-602-656-1 Sequence 1, Appli
13	36	55.4	4302	3	US-08-658-136-5 Sequence 3, Appli
14	36	55.4	4302	4	US-09-052-469-8 Sequence 8, Appli
15	36	55.4	4302	2	US-08-460-751-2 Sequence 2, Appli
16	36	55.4	4339	4	US-09-052-469-6 Sequence 6, Appli
17	35	53.8	140	3	US-08-930-894-4 Sequence 4, Appli
18	35	53.8	249	1	US-08-155-171B-32 Sequence 32, Appli
19	35	53.8	249	2	US-08-435-998-32 Sequence 32, Appli
20	35	53.8	512	1	US-08-173-508-4 Sequence 4, Appli
21	35	53.8	512	2	US-08-265-310-4 Sequence 4, Appli
22	35	53.8	512	3	US-08-951-742-4 Sequence 4, Appli
23	35	53.8	530	4	US-08-943-714-12 Sequence 12, Appli
24	35	53.8	956	3	US-08-772-270A-8 Sequence 8, Appli
25	35	53.8	1265	4	US-09-347-878-5 Sequence 5, Appli
26	35	53.8	1265	4	US-09-347-878-7 Sequence 7, Appli
27	34	52.3	79	4	US-08-936-165A-403 Sequence 403, App

28	34	52.3	93	2	US-08-341-843B-32 Sequence 32, Appli
29	34	52.3	93	2	US-08-427-497E-37 Sequence 37, Appli
30	34	52.3	238	4	US-09-605-858-35 Sequence 35, Appli
31	34	52.3	265	3	US-08-483-857-8 Sequence 8, Appli
32	34	52.3	270	4	US-09-134-001C-3942 Sequence 3942, Ap
33	34	52.3	307	4	US-09-134-001C-3238 Sequence 3238, Ap
34	34	52.3	311	4	US-09-564-805-230 Sequence 230, App
35	34	52.3	312	4	US-09-134-001C-5583 Sequence 5583, Ap
36	34	52.3	326	4	US-09-564-805-231 Sequence 231, App
37	34	52.3	448	4	US-09-627-650B-17 Sequence 17, Appli
38	34	52.3	448	4	US-09-436-063C-17 Sequence 17, Appli
39	34	52.3	475	4	US-09-627-650B-16 Sequence 16, Appli
40	34	52.3	475	4	US-09-436-063C-16 Sequence 16, Appli
41	34	52.3	487	4	US-09-627-650B-13 Sequence 13, Appli
42	34	52.3	487	4	US-09-627-650B-14 Sequence 14, Appli
43	34	52.3	487	4	US-09-436-063C-13 Sequence 13, Appli
44	34	52.3	487	4	US-09-436-063C-14 Sequence 14, Appli
45	34	52.3	506	4	US-09-627-650B-15 Sequence 15, Appli

ALIGNMENTS

```
RESULT 1
US-08-870-827-2
; Sequence 2, Application US/08870827
; Patent No. 5962297
;
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,827
FILING DATE: 06-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
TOPOLOGY: linear
FRAGMENT TYPE: internal fragment
US-08-870-827-2
Query Match 100.0%; Score 65; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SVFDGSDGTIVYQ 12
DB 1 SVFDGSDGTIVYQ 12
```

RESULT 2
US-09-317-179-2
Sequence 2, Application US/09317179
Patent No. 6383769
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/317,179
FILING DATE: 24-May-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/870,827
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
TOPOLOGY: linear
FRAGMENT TYPE: Internal fragment
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-317-179-2
Query Match 100.0%; Score 65; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SVFDGSGDTVYQ 12
Db 1 SVFDGSGDTVYQ 12
RESULT 3
US-08-870-827-3
Sequence 3, Application US/08870827
Patent No. 5962297
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,827
FILING DATE: 06-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-870-827-3
Query Match 100.0%; Score 65; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SVFDGSGDTVYQ 12
Db 201 SVFDGSGDTVYQ 212
RESULT 4
US-09-317-179-3
Sequence 3, Application US/09317179
Patent No. 6383769
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/317,179
FILING DATE: 24-May-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/870,827
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid

LOCATION: 905.952 /label
OTHER INFORMATION:
OTHER INFORMATION: /note

OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 188
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 238
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 318
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 437
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 453
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 474
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 501
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 571
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 913
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note="potential site of ASN-linked
OTHER INFORMATION: glycosylation"
US-08-408-093-6

Query Match 56.9% Score 37: DB 1: length 1018;
Best Local Similarity 77.8% Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 DGGDCTVYQ 12
|||||
Db 981 DGGDGVVSG 989

RESULT 7
US-08-408-420A-6
Sequence 6, Application US/08408420A
Patent No. 5731154
GENERAL INFORMATION:
APPLICANT: Reid, Robert A.
APPLICANT: Hemperly, John J.

TITLE OF INVENTION: Human Cell Adhesion Molecule and Nucleic
TITLE OF INVENTION: Acid Sequences
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
STREET: One Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/408,420A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fugitt, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-2630
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 45..94
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 138..191
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 243..290
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 332..371
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 416..464
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 506..563
FEATURE:
NAME/KEY: Domain
LOCATION: 604..657
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note="conserved core of fibronectin type
OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 707..760
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note="conserved core of fibronectin type
OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 809..857
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note="conserved core of fibronectin type
OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 905..952
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note="conserved core of fibronectin type

OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 188
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
FEATURE:
NAME/KEY: Modified-site
LOCATION: 238
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
FEATURE:
NAME/KEY: Modified-site
LOCATION: 318
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
FEATURE:
NAME/KEY: Modified-site
LOCATION: 437
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
FEATURE:
NAME/KEY: Modified-site
LOCATION: 453
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
FEATURE:
NAME/KEY: Modified-site
LOCATION: 474
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
FEATURE:
NAME/KEY: Modified-site
LOCATION: 501
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
FEATURE:
NAME/KEY: Modified-site
LOCATION: 571
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
FEATURE:
NAME/KEY: Modified-site
LOCATION: 913
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: /label= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
US-08-408-420A-6

Query Match 56.9%; Score 37; DB 1; Length 1018;
Best Local Similarity 77.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 DGGDGTYYQ 12
DB 981 DGGDGTYYQ 989

RESULT 8
US-08-714-901-6
Sequence 6, Application US/08714901
Patent No. 5739289
GENERAL INFORMATION:
APPLICANT: Reid, Robert A.
APPLICANT: Hemperly, John J.

TITLE OF INVENTION: Human Cell Adhesion Molecule and Nucleic
TITLE OF INVENTION: Acid Sequences
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson
ADDRESSEE: and Company
STREET: One Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: USA
ZIP: 07417
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,901
FILING DATE: 17-SEP-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/408,093
FILING DATE: 21-MAR-1995
APPLICATION NUMBER: US/08/040,741
FILING DATE: 26 MAR 1993
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-2630
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo. sapiens
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 45..94
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 138..191
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 243..290
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 332..371
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 416..464
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 506..563
FEATURE:
NAME/KEY: Domain
LOCATION: 604..657
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 707..760
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIT-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 809..857
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIT-like repeat"

```
FEATURE:
NAME/KEY: Domain
LOCATION: 905..952
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: tit-like repeat"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 188
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 238
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 318
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 437
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 453
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 474
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 501
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 571
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 913
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
US-08-714-901-6

Query Match          56.9%; Score 37; DB 1; Length 1018;
Best Local Similarity 77.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

RESULT 9
US-08-452-052-2

```
; Sequence 2, Application US/08452052
; Patent No. 5766922
; GENERAL INFORMATION:
; APPLICANT: Peles, Elior
; TITLE OF INVENTION: FUNCTIONAL LIGANDS FOR THE AXONAL CELL
; TITLE OF INVENTION: RECOGNITION MOLECULE CONTACTIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,052
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7683-075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-452-052-2
```

```
Query Match          56.9%; Score 37; DB 1; Length 1018;
Best Local Similarity 77.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY      4 DGGDGTVO 12
      11111111
DB      980 DGGDGTVO 988

RESULT 10
US-08-040-741-6
; Sequence 6, Application US/08040741
; Patent No. 6017695
; GENERAL INFORMATION:
; APPLICANT: Reid, Robert A.
; APPLICANT: Hemperly, John J.
; TITLE OF INVENTION: Human Cell Adhesion Molecule and Nucleic
; TITLE OF INVENTION: Acid Sequences
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson
; ADDRESSEE: and Company
; STREET: One Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: USA
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/040,741
FILING DATE: 19930326
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: FUGIT, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-2630
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1018 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 45..94
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 138..191
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 243..290
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 332..371
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 416..464
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 506..563
FEATURE:
NAME/KEY: Domain
LOCATION: 604..657
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIR-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 707..760
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIR-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 809..857
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIR-like repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 905..952
OTHER INFORMATION: /label= FLR
OTHER INFORMATION: /note= "conserved core of fibronectin type
OTHER INFORMATION: IIR-like repeat"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 188
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 238
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
OTHER INFORMATION: glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 318
OTHER INFORMATION: /label= ASN-glycos

OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 437
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 453
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 474
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 501
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 571
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 913
OTHER INFORMATION: /label= ASN-glycos
OTHER INFORMATION: /note= "potential site of ASN-linked
glycosylation"
OTHER INFORMATION: glycosylation"
US-08-040-741-6
Query Match 56.9%; Score 37; DB 3; Length 1018;
Best Local Similarity 77.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Caps 0;
QY 4 DGGDGYQYQ 12
DB 981 DGGDGVVQ 989
RESULT 11
US-08-289-709-1
Sequence 1, Application US/08289709
Patent No. 5523224
GENERAL INFORMATION:
APPLICANT: Burtcher, Helmut; Lang, Gunter; Popp, Friedrich
TITLE OF INVENTION: Recombinant D-Hydantoinase, A Process For the
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,709
FILING DATE: 12-AUGUST-1994

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P 43 28 829.4
FILING DATE: 27-AUGUST-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 553222aman D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: BOER 1041
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-289-709-1

Query Match 55.4%; Score 36; DB 1; Length 460;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 SVFDGDSGVYQ 12
:11 1111:
Db 157 NVFQADDGTVLR 168

RESULT 12
US-08-602-656-1
Sequence 1, Application US/08602656
Patent No. 5679571
GENERAL INFORMATION:
APPLICANT: Bartscher, Helmut; Lang, Gunter; Popp, Friedrich
TITLE OF INVENTION: Recombinant D-Hydantoinase, A Process For the
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,656
FILING DATE: 16-FEB-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/289,709
FILING DATE: 12-AUGUST-1994
APPLICATION NUMBER: P 43 28 829.4
FILING DATE: 27-AUGUST-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5679571man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: BOER 1041
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-602-656-1

Query Match 55.4%; Score 36; DB 1; Length 460;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 SVFDGDSGVYQ 12
:11 1111:
Db 157 NVFQADDGTVLR 168

RESULT 13
US-08-658-136-5
Sequence 5, Application US/08658136
Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GACKOWSKI, GREGORY
APPLICANT: OLAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN4-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-5415
TELEFAX: 508-872-8400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-658-136-5

Query Match 55.4%; Score 36; DB 3; Length 4302;
Best Local Similarity 87.5%; Pred. No. 2.1e+03;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 FDGDSGVY 10
:11 11111:
Db 1245 FDMGDSGVY 1252

RESULT 14
US-09-052-469-8
Sequence 8, Application US/09052469
Patent No. 6380360
GENERAL INFORMATION:
APPLICANT: Harris et al.
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Wilcoff, Ltd.
STREET: One Financial Center
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk, 3.50 inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/052,469
FILING DATE: Concurrently herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/422,582
FILING DATE: 14-April-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9507766.5
FILING DATE: 13-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9411900.5
FILING DATE: 14-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB PCT/GB94/02822
FILING DATE: 23-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9326470.3
FILING DATE: 24-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Ph.D., Kathleen M.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 3265/74165
TELEPHONE: (617) 345-9100
TELEFAX: (617) 345-9111
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-052-469-8
Query Match 55.4%; Score 36; DB 4; Length 4302;
Best Local Similarity 87.5%; Pred. NO. 2.1e+03;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 FDGDDGTV 10
Db 1245 FDMGDDTV 1252
RESULT 15
US-08-460-751-2
Sequence 2, Application US/08460751
Patent No. 5891628
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen
APPLICANT: Schneider, Michael
APPLICANT: Glucksmann, Sandra
TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
DISEASE GENE, DIAGNOSTICS AND TREATMENT
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,751
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/413,580
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7638-005
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4303 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-460-751-2
Query Match 55.4%; Score 36; DB 2; Length 4303;
Best Local Similarity 87.5%; Pred. NO. 2.1e+03;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Search completed: June 16, 2003, 12:11:32
Job time : 2.08197 secs


```

XX Chaeen H, Kubota M, Tsusaki K:
PI WPI: 1998-034976/04.
DR N-PSDB; AAV17621.
XX
XX Bacillus beta-fructofuranosidase enzyme - useful for producing
PT fructo:furanosylated saccharide(s) or alcohol(s)
PS
XX
XX Claim 1: Pages 23-25; 32pp; English.

This is beta-fructofuranosidase from Bacillus sp. V230, an enzyme that
CC has an optimum temperature of 50 deg. C at pH 6. It catalyses
CC fructofuranosyl transfer reactions from a fructofuranosyl donor to a
CC fructofuranosyl acceptor where the donor is sucrose, raffinose or erlose
CC and the acceptor is selected from alcohols, sugar alcohols and
CC saccharides having no beta-fructofuranosidic linkages,
CC especially where the reaction product is xylosyl fructoside, erlose,
CC isomaltosyl fructoside, lactosucrose or fructosyltrehalose.
CC Such reaction products have a satisfactory taste and sweetness, a
CC moderate viscosity and humectancy, an effective anticariogenic activity,
CC growth promoting activity for Bifidobacteria, mineral-absorption
CC promoting activity. The products can be used to improve the tastes and
CC textures of food products, cosmetics and pharmaceuticals and are useful
CC as sweeteners in the food and pharmaceutical industries.
CC
XX
XX Sequence 487 AA:
SQ
XX
XX Query Match 100.0%; Score 2454; DB 19; Length 487;
XX Best Local Similarity 100.0%; Pred. No. 7.6e-197;
XX Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNSGDKYKDYGFANHTTRADMLKIPGOQNSPOKRYVQFNASAKINDSAKGYDKSGNLIDL 60
DB 33 MNSGDKYKDYGFANHTTRADMLKIPGOQNSPOKRYVQFNASAKINDSAKGYDKSGNLIDL 92
OY 61 DWDSWPLQONADGTANHYGHYIVSALAGDPKNSDPTLHLYOKYVGTSDSKWNAQRY 120
DB 93 DWDSWPLQONADGTANHYGHYIVSALAGDPKNSDPTLHLYOKYVGTSDSKWNAQRY 152
OY 121 FEDMDKFPNDPRLKYTOEWSGSAITLTKDGOVRLFTYDYSGNPDDGGAGNOIITSAO 180
DB 153 FEDMDKFPNDPRLKYTOEWSGSAITLTKDGOVRLFTYDYSGNPDDGGAGNOIITSAO 212
OY 181 VNLSPDAATLKVDGVSDFHKSVPFDGSGTIVYONIOQFIDECKWISGDHNTLDRPHYVEDK 240
DB 213 VNLSPDAATLKVDGVSDFHKSVPFDGSGTIVYONIOQFIDECKWISGDHNTLDRPHYVEDK 272
OY 241 GHKYLVEFANVTGTDGTYGODGSFNKKAYGSDVFFONEKNKLLQSPKKQIASLANGALG 300
DB 273 GHKYLVEFANVTGTDGTYGODGSFNKKAYGSDVFFONEKNKLLQSPKKQIASLANGALG 332
OY 301 IVELADDTYKSVKMPVLAASNTVADEVERANFEKNNKMYLFTDSRGSKMTSDGINDKV 360
DB 333 IVELADDTYKSVKMPVLAASNTVADEVERANFEKNNKMYLFTDSRGSKMTSDGINDKV 392
OY 361 YMLGGGDSLNGPNNPINETGVLVLMNLDPAADLTHTYSHCGIPHEGNNVVLTSYMTNRG 420
DB 393 YMLGGGDSLNGPNNPINETGVLVLMNLDPAADLTHTYSHCGIPHEGNNVVLTSYMTNRG 452
OY 421 FYPEHHSHLRDKLGVNIGSDTSGGESSGGOGFP 455
DB 453 FYPEHHSHLRDKLGVNIGSDTSGGESSGGOGFP 487

```

```

DE Novel human diagnostic protein #4298.
XX
XX Human: Chromosome mapping; gene mapping; gene therapy; forensic:
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HSE-) HSEO INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI: 2001-639362/73.
XX
XX N-PSDB; AAS68494.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 34666; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 527 AA:
SQ
XX
XX Query Match 63.3%; Score 1553; DB 22; Length 527;
XX Best Local Similarity 65.3%; Pred. No. 2.5e-121;
XX Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;

2 NSGDKYKDYGFANHTTRADMLKIPGOQNSPOKRYVQFNASAKINDSAKGYDKSGNLIDL 61
DB 88 NOKAKETTYGVSHTRHMDLIPKQONEKYQVDFPDOSTKNIESAG-----LD 138
OY 62 VMDSWPLQONADGTANHYGHYIVSALAGDPKNSDPTLHLYOKYVGTSDSKWNAQRY 121
DB 139 VMDSWPLQONADGTANHYGHYIVSALAGDPKNSDPTLHLYOKYVGTSDSKWNAQRY 198
OY 122 EDMKFPNDPRLKYTOEWSGSAITLTKDGOVRLFTYDYSGNPDDGGAGNOIITSAO 181
DB 199 KDSKFDANDPTLKDQTOEWSGSAITLTKDGOVRLFTYDYSGNPDDGGAGNOIITSAO 252
OY 182 NLSQPDAAITLKVDGVSDFHKSVPFDGSGTIVYONIOQFIDECKWISGDHNTLDRPHYVEDK 241
DB 253 NVKRSK-DTLKINGVEDHKITFD-GDCKTYQNVQOQFIDECKWISGDHNTLDRPHYVEDK 310

```



```

QY 242 HKYVFEANTGTTDYGQDSFNNKAYGSDVFFQNEKNKLLQSPKKQIASLANGALGI 301
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC 311 HKYVFEANTGTTGNGYGGESLFNKAYGGGTFNFRKESQKLOQSAKKRDAELANGALGI 370
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 302 VELADDTYVKSVMKPLVANSNTVADEVERANIFKMNKKWYLFETDSRSGSKMTSGINDKDYV 361
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 371 IELNNDYTLKKYMKPLITNTVTDEIERANVFKMNGKWLFLFDSRSGSKMTIDGINSNDIY 430
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 362 MLGPGDSSLNGPHNFINETGLVLMNLDPADLTHYSHCGIPHPGNNVVLTSYMTNRGF 421
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 431 MLGYVNSLTCGYKPLNKTGLVLMGLDLPNDVFTYSHFAVPQAKGNVVTISYMTNRGF 490
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 422 YPEHSHLRDKLGVNFKGSDTSNGENS-SGQGO 453
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 491 FEDKATFAPSFMLNFKGKTSVKNLSILEQGO 523
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 3
ABG06230
ID ABG06230 standard; Protein; 644 AA.
XX
AC ABG06230;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #6221.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
  food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PT 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR MPI: 2001-639362/73.
XX
N-PSDB: AAS70417.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
  diagnostics, forensics, gene mapping, identification of mutations
  responsible for genetic disorders or other traits and to assess
  biodiversity
XX
PS Claim 20; SEQ ID No 36589; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
  polypeptide (II) sequences. (I) is useful as hybridisation probes,
  polymerase chain reaction (PCR) primers, oligomers, and for chromosome
  and gene mapping, and in recombinant production of (II). The
  polynucleotides are also used in diagnostics as expressed sequence tags
  for identifying expressed genes. (I) is useful in gene therapy techniques
  to restore normal activity of (II) or to treat disease states involving
  (II). (II) is useful for generating antibodies against it, detecting or
  quantitating a polypeptide in tissue, as molecular weight markers and as
  a food supplement. (II) and its binding partners are useful in medical
  imaging of sites expressing (II). (I) and (II) are useful for treating
  disorders involving aberrant protein expression or biological activity.
  CC The polypeptide and polynucleotide sequences have applications in
  diagnostics, forensics, gene mapping, identification of mutations
  responsible for genetic disorders or other traits to assess biodiversity
  and to produce other types of data and products dependent on DNA and
  amino acid sequences. ABG00010-ABG30377 represent novel human

```

CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

```

SQ Sequence 644 AA;
Query Match 63.3%; Score 1553; DB 22; Length 644;
Best Local Similarity 65.3%; Pred. No. 3.3e-121;
Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;

QY 2 NSGDYKEDYGFAHIFRADMLKIPGQNSPOFVKVPOFNASAKINIDSAGYDKSGNLIDD 61
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 205 NOKAYKETGYVSHITFHDMLQIPKQONKEKYVPPQDQSTIKNIESAKG-----LD 255
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 VWDSPFLQADDTAANYHGYHIVSALAGDPKNSDPTPLFLFYOKVGDTSIDSKNAGRV 121
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 256 VWDSPFLQADDTVAEYNGYHVAFALAGSPKADDTSIYMFYOKYGDNSIDSKNAGRV 315
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 122 EDMDKFVPNDPYLKYQTOEMSGSATLTTRKQYVLFYTDYSGNPEDEGTAGNOIISTAOV 181
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 316 KOSDKFDANDPILKQDOTQEMSGSATFTSDGKIRLFYTDYSGK-----HYGQSILTTAOV 369
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 182 NLSQPDATLKYDGYSDHKSVEFDGDTGYVQNIQFIDECKWISGDNHTLRDPHYVEDKG 241
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 370 NVSKSD-DTLKINGYEDHKTFED-GDGKTYQNVQCFIDEGNVTSGDNHTLRDPHYVEDKG 427
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 242 HKYVFEANTGTTDYGQDSFNNKAYGSDVFFQNEKNKLLQSPKKQIASLANGALGI 301
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 428 HKYVFEANTGTTGNGYGGESLFNKAYGGGTFNFRKESQKLOQSAKKRDAELANGALGI 487
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 302 VELADDTYVKSVMKPLVANSNTVADEVERANIFKMNKKWYLFETDSRSGSKMTSGINDKDYV 361
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 488 IELNNDYTLKKYMKPLITNTVTDEIERANVFKMNGKWLFLFDSRSGSKMTIDGINSNDIY 547
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 362 MLGPGDSSLNGPHNFINETGLVLMNLDPADLTHYSHCGIPHPGNNVVLTSYMTNRGF 421
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 548 MLGYVNSLTCGYKPLNKTGLVLMGLDLPNDVFTYSHFAVPQAKGNVVTISYMTNRGF 607
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 422 YPEHSHLRDKLGVNFKGSDTSNGENS-SGQGO 453
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 608 FEDKATFAPSFMLNFKGKTSVKNLSILEQGO 640
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
ABG22589
ID ABG22589 standard; Protein; 736 AA.
XX
AC ABG22589;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #22580.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
  food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PT 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR MPI: 2001-639362/73.

```

DR N-PSDB; AAS86776.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20: SEQ ID No 52948; 103pp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 736 AA:
 SQ
 Query Match 63.3%; Score 1553; DB 22: Length 736;
 Best Local Similarity 65.3%; Pred. No. 4, 1e-121;
 Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;
 QY 2 NSGDYKEDYGFHITRADMLKIPGOONSPOFKYPOFNASAIKNSAKYDKSGLIDDD 61
 DB 297 NOKAYKETYGVSHTIRHDMQIPKQOONKRYQVPOFDOSTIKIESAKG-----LD 347
 QY 62 VWDSPFLQADGTAAANYHGIVTSALAGDPKNSDDPPLHLFYQKVGDTSIDSKNAGRYF 121
 DB 348 VWDSPFLQADGTAAANYHGIVTSALAGDPKNSDDPPLHLFYQKVGDTSIDSKNAGRYF 407
 QY 122 EDMKFEVNDPLKTYOTQDWSGSATLTKDGOVRLFTYTDYSGNPEDEGTAGNOIISTAOV 181
 DB 408 KDSDKRDANDPLIKDQTOEMSGSATFTSGKIRLEFTYDYSK-----HYGKSLTTAGV 461
 QY 182 NLSQPPAATLKYDGVSDHKSVEEDGGGTYOQNFIDSGKWTSGDNHLLRDPHYEDKG 241
 DB 462 NVSKSD-DPLKINGVEDHKTIFD-GDGKTYOANVOQFIDEGNYSGNHNLRLRPHYVEDKG 519
 QY 242 HNYLVEANTGTGTDGVOGDSFNKNKAYYGGSDVFPENKKNKILQSKKQIASLANALAI 301
 DB 520 HNYLVEANTGTGTDGVOGDSFNKNKAYYGGSDVFPENKKNKILQSKKQIASLANALAI 579
 QY 302 VELADDTYVSKYKPLVASNTVADEVERANIFKMNKMYLFTDNRSGSKMTSDGINDKDY 361
 DB 580 IELNDYTLTKYMKPLTISNTYTDEITERANVFKNMGKMYLFTDNRSGSKMTSDGINDKDY 639
 QY 362 MGPFGDSLNGPHNPINETGVLNNMLDPAIDLTHYSHGCIIPHPEGNNVLTSTYMTNRCF 421
 DB 640 MGYVSNLSLTGYPKPLNTKGLVLOMGLDENDVTFTSHFAPVQAKNNVITSYMTNRCF 699
 QY 422 YPEHSHLADKLGVNIGKSDTSGGENS-SGOGO 453
 DB 700 FEDKATGPAFSPFLMNKMGKMTSVVKNLSLEOGO 732

RESULT 5
 ABG05590
 ID ABG05590 standard; Protein; 774 AA.

XX
 AC ABG05590;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #5581.
 XX
 KM Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YF;
 XX
 DR WPI: 2001-639362/73.
 DR N-PSDB: AAS69777.
 XX
 XX Claim 20: SEQ ID No 35949; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 774 AA:
 SQ
 Query Match 63.3%; Score 1553; DB 22: Length 774;
 Best Local Similarity 65.3%; Pred. No. 4, 4e-121;
 Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;
 QY 2 NSGDYKEDYGFHITRADMLKIPGOONSPOFKYPOFNASAIKNSAKYDKSGLIDDD 61
 DB 35 NOKAYKETYGVSHTIRHDMQIPKQOONKRYQVPOFDOSTIKIESAKG-----LD 85
 QY 62 VWDSPFLQADGTAAANYHGIVTSALAGDPKNSDDPPLHLFYQKVGDTSIDSKNAGRYF 121
 DB 86 VWDSPFLQADGTAAANYHGIVTSALAGDPKNSDDPPLHLFYQKVGDTSIDSKNAGRYF 145
 QY 122 EDMKFEVNDPLKTYOTQDWSGSATLTKDGOVRLFTYTDYSGNPEDEGTAGNOIISTAOV 181

Db 146 KDSKFDANDPILKDOTQEMSGSATFTSDGKIRLFYTDYSGK -----HYGQSILTTAAQV 199
 QY 182 NLSQPDATLTKVDGYSVDRKSVFDDGDTYVONIQQFIDEGKMWISGDNHTLRDPHYVEDKG 241
 Db 200 NVSKSD-DTLKINGVEDHKTIFD-GDGKTYQVQVOQFIDEGANTSGDNHTLRDPHYVEDKG 257
 QY 242 HKYLVFEANTGTGGDYGDSQFNNKAYVGGSDVFPQNEKNKLLQSPKQIASLANGALGI 301
 Db 258 HKYLVFEANTGTENGQGEESLFNKAYVGGGTFNFRKESQKIQQSARKKDAELANGALGI 317
 QY 302 VELADDTYVKSVMKPLVASNTVADVEERANIFKMNKKWYLFETDSGSKMTSGINDKDYV 361
 Db 318 IELNDYTLTKKVMKPLITSNTVTFDEIERANVFKMGKWLFTFDSGSKMTIDGINSNDIY 377
 QY 362 MCGPGDGLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHPGNNVLTSTYTNKGF 421
 Db 378 MCGVNSLTPGPKYKPLKNTGLVLMGDLDPNDVTFYSHFAVQAQKNNVITSYTNKGF 437
 QY 422 YPEHSHLRDKLGVNIKSGDTSGENS-SGQGG 453
 Db 438 FEDKKATFAPSPFLMNKGNKTSYVKNLSILEGG 470

RESULT 6

ABG11754 standard; Protein: 774 AA.

ABG11754;

18-FEB-2002 (first entry)

Novel human diagnostic protein #11745.

Human: chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI: 2001-639362/73.

N-PSDB; AAS75941.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity

Claim 20; SEQ ID No 42113; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG3037 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 774 AA:

Query Match 63.3%; Score 1553; DB 22; Length 774;

Best Local Similarity 65.3%; Pred. No. 4.4e-121;

Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;

QY 2 NSGDKEDYGFPHITRADMLKIPGQONSQFQVPPQPMASATINDSAKYDKSGNLIDD 61
 Db 35 NOKAVKETYGVSHITRHDMLQIPKQOQNEKYVPOFDOSTIKNISAKG-----ID 85
 QY 62 VMDSWPLOWADGTVAENYGHVYSALAGDPKNSDPTPLHLEFYQKGDPSIDSMKNAGRVF 121
 Db 86 VMDSWPLOWADGTVAENYGHVYFALAGSPKADPTSTYFYQKVDNSIDSMKNAGRVF 145
 QY 122 EDMKFPVNDPYLKYQTOEMSGSATLTKDQVRLFYTDYSGNPDGDTGAGNQIISTAAQV 181
 Db 146 KDSKFDANDPILKDOTQEMSGSATFTSDGKIRLFYTDYSGK -----HYGQSILTTAAQV 199
 QY 182 NLSQPDATLTKVDGYSVDRKSVFDDGDTYVONIQQFIDEGKMWISGDNHTLRDPHYVEDKG 241
 Db 200 NVSKSD-DTLKINGVEDHKTIFD-GDGKTYQVQVOQFIDEGANTSGDNHTLRDPHYVEDKG 257
 QY 242 HKYLVFEANTGTGGDYGDSQFNNKAYVGGSDVFPQNEKNKLLQSPKQIASLANGALGI 301
 Db 258 HKYLVFEANTGTENGQGEESLFNKAYVGGGTFNFRKESQKIQQSARKKDAELANGALGI 317
 QY 302 VELADDTYVKSVMKPLVASNTVADVEERANIFKMNKKWYLFETDSGSKMTSGINDKDYV 361
 Db 318 IELNDYTLTKKVMKPLITSNTVTFDEIERANVFKMGKWLFTFDSGSKMTIDGINSNDIY 377
 QY 362 MCGPGDGLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHPGNNVLTSTYTNKGF 421
 Db 378 MCGVNSLTPGPKYKPLKNTGLVLMGDLDPNDVTFYSHFAVQAQKNNVITSYTNKGF 437
 QY 422 YPEHSHLRDKLGVNIKSGDTSGENS-SGQGG 453
 Db 438 FEDKKATFAPSPFLMNKGNKTSYVKNLSILEGG 470

RESULT 7

ABG04946 standard; Protein: 789 AA.

ABG04946;

13-FEB-2002 (first entry)

Novel human diagnostic protein #4937.

Human: chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

|||||
Db 401 VWDSPWPLQADGTVAEYNGYHVAFALAGSPKADADTSTYMEFYQKVGDNISIDSMKNAGRVE 460
QY 122 EDMKFVENDPYLKYQTOEWSGSAATLTKDGOVRLFTYTDGSPNEDGGTGAGAGQIISTAYOV 181
Db 461 KDSKFEFANDPILDKQTOEWSGSAATLTKDGOVRLFTYTDGSPNEDGGTGAGAGQIISTAYOV 514
QY 182 NLSQPDAAATLKYDVSVDHKSVDGSDGTGYVONIOQFIDEGKWSISGDNHTLDPHYVEDKG 241
Db 515 NVKSSD-DTLKINGVEDHKITFD-GDGKTYQNVQOFIDEGVNTSGDNHTLDPHYVEDKG 572
QY 242 HKYLFVFNANTGTGGYQGDOSFNKNKAYGSDVFFQNKKNLQSPKQIASLANGALGI 301
Db 573 HKYLFVFNANTGTGGYQGDOSFNKNKAYGSDVFFQNKKNLQSPKQIASLANGALGI 632
QY 302 VELADDTYKSVKPLVASNTVADEVERANIFKNNKMYLFTDSGSKMTSDGINDKDYV 361
Db 633 IELNNDYTLKKYMKPLITSNTVTDEIERANVFKMNGKMYLFTDSGSKMTSDGINDKDYV 692
QY 362 MLGPGGDSLNGPHNINTEGLVLMNLDPADLTHYSHGCIPIHPEGNNVLTSTYTNNGF 421
Db 693 MLGYVNSLTGPKPLKNTGLVLQMGDLPNDVTFTYSHFAVPAQAKGNVVTSTYTNNGF 752
QY 422 YPEHSHLRDKLGVNIKSGDTSKGENS-SGQGO 453
Db 753 FEDKKATFAPSPFLMNKIKGNTSVVKNSTILEQGO 785

RESULT 9
ABG28407
ID ABG28407 standard; Protein: 789 AA.
AC ABG28407;
XX
XX 18-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #28398.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PE
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX
XX 23-AUG-2000; 2000US-0649167.
PS
XX
XX (HYSE-) HYSEQ INC.
PI
XX
XX Drmanac RT, Liu C, Tang YT;
DR
XX
XX WPI: 2001-639362/73.
DR
XX
XX N-PSDB; AAS92594.
PT
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
PS
XX
XX Claim 20: SEQ ID NO 58766; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG3037 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC publication, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
SQ
Sequence 789 AA:
Query Match 63.3%; Score 1553; DB 22; Length 789;
Best Local Similarity 65.3%; Pred. No. 4,5e-121;
Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;
QY 2 NSGDYKEDYGFATHRADMLKIPGOONSPOEFVPOFNASAIKNIDSAGYDGSGLDLD 61
Db 350 NOKAKETTYGVSHIRHMLQIPKQOQMEKYVPPFDOSTINKIESAGK-----LD 400
QY 62 VWDSPWPLQADGTVAEYNGYHVAFALAGSPKADADTSTYMEFYQKVGDNISIDSMKNAGRVE 121
Db 401 VWDSPWPLQADGTVAEYNGYHVAFALAGSPKADADTSTYMEFYQKVGDNISIDSMKNAGRVE 460
QY 122 EDMKFVENDPYLKYQTOEWSGSAATLTKDGOVRLFTYTDGSPNEDGGTGAGAGQIISTAYOV 181
Db 461 KDSKFEFANDPILDKQTOEWSGSAATLTKDGOVRLFTYTDGSPNEDGGTGAGAGQIISTAYOV 514
QY 182 NLSQPDAAATLKYDVSVDHKSVDGSDGTGYVONIOQFIDEGKWSISGDNHTLDPHYVEDKG 241
Db 515 NVKSSD-DTLKINGVEDHKITFD-GDGKTYQNVQOFIDEGVNTSGDNHTLDPHYVEDKG 572
QY 242 HKYLFVFNANTGTGGYQGDOSFNKNKAYGSDVFFQNKKNLQSPKQIASLANGALGI 301
Db 573 HKYLFVFNANTGTGGYQGDOSFNKNKAYGSDVFFQNKKNLQSPKQIASLANGALGI 632
QY 302 VELADDTYKSVKPLVASNTVADEVERANIFKNNKMYLFTDSGSKMTSDGINDKDYV 361
Db 633 IELNNDYTLKKYMKPLITSNTVTDEIERANVFKMNGKMYLFTDSGSKMTSDGINDKDYV 692
QY 362 MLGPGGDSLNGPHNINTEGLVLMNLDPADLTHYSHGCIPIHPEGNNVLTSTYTNNGF 421
Db 693 MLGYVNSLTGPKPLKNTGLVLQMGDLPNDVTFTYSHFAVPAQAKGNVVTSTYTNNGF 752
QY 422 YPEHSHLRDKLGVNIKSGDTSKGENS-SGQGO 453
Db 753 FEDKKATFAPSPFLMNKIKGNTSVVKNSTILEQGO 785

RESULT 10
ABG25769
ID ABG25769 standard; Protein: 823 AA.
AC ABG25769;
XX
XX 18-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #25760.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX

```

PF 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HSE-) HSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB; AAS89956.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 56128; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 823 AA:
SQ
Query Match 63.3%; Score 1553; DB 22; Length 823;
Best Local Similarity 65.3%; Pred. No. 4.8e-121;
Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;
OY 2 NSGDYKEDYGFARHTRADMLKIPGOQNSPQFVPPFPMASATINIDSAGYDKSGMLDLD 61
DB 384 NOKAVKETTYGVSHIRHMLQIPKQONEKYQVPOFDOSTIKINISAKG-----LD 434
OY 62 VWDSPPLONADGTANVGHYHVSALADPKNSDPTPLHLEFOKQGDNSIDSMKNAGRYE 121
DB 435 VWDSPPLONADGTAEVNGYHVFALASPKDADDTSTMFQKQGDNSIDSMKNAGRYE 494
OY 122 EDMDKFVNDPYLYKQTOEWSSATLFRKDGQVRLFYTTYSGNPEDGTCAGNQIISTAYQV 181
DB 495 KDSDFEANDPILIKQTOEWSSATFTSDGKIRLRYTYSGR-----HYGQSLTTFAYQV 548
OY 182 NLSQPDATLKVADGYSDKHSVFDGSDGYVYQNIQFIDBGKGISDNDHLRDPHYVEDKG 241
DB 549 NVSKSD-DLTKINGVEDHKITFD-GDGKTYQVVOQFIDEGNTSDNHTLRDPHYVEDKG 606
OY 242 HKYLVFEANTGTGTDYOGDSFNNKAYYGGSDVFQNKKNKLQSPKQIATSLANGALGI 301
DB 607 HKYLVFEANTGTENGYOGESEFNFKAATYGGGTNFRKKSOKIKQASAKRDELALGALGI 666
OY 302 VELADDTYKSVMKPLVASNTVADEVERANIFKNNKMYLFTDSGSKMTSDGINDKDYV 361
DB 667 IELNNDYLYKVMKPLVSNVTYTBIEIRANVFKMKGMYLFTDSGSKMTSDGINSNDIY 726
OY 362 MLGPGGDSLGNPHNFINETGLVANNLDPADLTHYTSKCGIPHPGNNVLTSTYWTNKG 421

```

```

DB 727 MLGYVNSLTPGEKPLNKTLGLQMLDLPNDVFTTYSHPFAVPOAKGNVVTSTWNRGF 786
OY 422 YPEHSHLRDKLGAVNIKSDPTSGCENS-SGGOQ 453
DB 787 FEDKATRAPSPFLMNKIKNTSVKNSILEQO 819
RESULT 11
ABG04500
ID ABG04500 standard; Protein; 855 AA.
XX
XX ABG04500;
XX
XX 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #4491.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HSE-) HSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB; AAS68687.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 34859; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 855 AA:
SQ
Query Match 63.3%; Score 1553; DB 22; Length 855;
Best Local Similarity 65.3%; Pred. No. 5.1e-121;
Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;

```


XX WO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI: 2001-639362/73.
 DR N-PSDB; AAS71354.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

PS Claim 20; SEQ ID NO 37526; 103pp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 893 AA;
 XX
 XX

Query Match 63.3%; Score 1553; DB 22; Length 893;
 Best Local Similarity 65.3%; Pred. No. 5,5e-121;
 Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;

OY 2 NSGDKEYGFAHTRADMLKIPGOONSPOFVPOFNASAIKINDSAGYKSSGLDLD 61
 DB 454 NOKAYKEYYGVSHTRHMLQIPKQONKTYOVPOFDSTIKNISAG-----LD 504
 OY 62 WDSWPLONADGTAAHYGHYVLSALACDPKNSDTPHLFYOXKGDTSIDSMKNAGRYE 121
 DB 505 WDSWPLONADGTAEYGHYVLSALACDPKNSDTPHLFYOXKGDTSIDSMKNAGRYE 564
 OY 122 EDMDFVNDPFLKQTOEWSGSATLTFDGOVRLFTYDSGNPEDGGTGAGNQITSAOV 181
 DB 565 KDSDFDNDPILKQTOEWSGSATLTFDGOVRLFTYDSGNPEDGGTGAGNQITSAOV 618
 OY 182 NLSQDPAATLKVDSGSHKSVFDDGSGTVYONIQOFTDEGKWSISDNHTLRPHYVEDKG 241
 DB 619 NVSSSD-DLTKINGEDHKTTFD-GDGKTYOVQOQFIDEGNTSGDNHTLRPHYVEDKG 676
 OY 242 HKYLVFEANTGTGDOGSDFNNKAYYGSDFVQONKKNLQPKKOIASLANGALGI 301
 DB 677 HKYLVFEANTGTGDOGSDFNNKAYYGSDFVQONKKNLQPKKOIASLANGALGI 736
 OY 302 VELADYTVKSVKPLVASNTVADEVERANIFKNNKMYLFTDSGSKMTSDGINDKDVY 361

DB 737 IELNDYTLKKVMKDLINSNTVTDEIERANVEKNMGKXYLFETDSGSKMTIDGINSNDIY 796
 OY 362 MLGPGGDSLNGPHNININTEGLVANNLDPADLTHRYSHCGIPHEGNVNVLTSTWTFNGF 421
 DB 797 MLGYVNSLGTGPKYKPLNKTGLVLOMGIDLPNDVTFYSHFAVPAQKGNVNVITSTWTFNGF 856
 OY 422 YPEHSHLRDRKLGYNKSGSDTSGENS-SGOGQ 453
 DB 857 FEDKATAPSPFLMNKIKNTSYVKNLSILEGQ 889

RESULT 14
 ABG02461
 ID ABG02461 standard; Protein; 913 AA.
 XX
 XX ABG02461;
 AC
 XX 13-FEB-2002 (first entry)
 DT
 XX
 XX Novel human diagnostic protein #2452.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI: 2001-639362/73.
 XX N-PSDB; AAS66648.
 DR

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

PS Claim 20; SEQ ID NO 32820; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 913 AA;
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:04:38 ; Search time 41.0246 Seconds

(Without alignments)
326.327 Million cell updates/sec

Title: US-09-986-682B-3

Perfect score: 2454
Sequence: 1 MNSGDYKEDYGFHAITRADM.....NIKGSITSGGENSSGQGF 455Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA.*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCRTS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/Backfilest1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2454	100.0	455	2	US-08-870-827-3
2	2454	100.0	455	4	US-09-317-179-3
3	281.5	11.5	543	1	US-08-362-232-2
4	281.5	11.5	543	1	US-08-814-196-2
5	278.5	11.3	578	4	US-09-503-172A-2
6	255	10.4	415	1	US-08-381-936-2
7	255	10.4	415	3	US-08-943-374-2
8	128	5.2	1475	3	US-09-007-999-2
9	128	5.2	1475	4	US-09-210-361-2
10	128	5.2	1475	4	US-09-740-274-2
11	116.5	4.7	1848	4	US-08-296-791-6
12	116.5	4.7	1848	5	PCT-US95-10661A-6
13	116	4.7	513	4	US-09-134-001C-4469
14	115.5	4.7	1545	4	US-08-296-791-4
15	115.5	4.7	1545	5	PCT-US95-10661A-4
16	115	4.7	931	4	US-08-624-655A-2
17	115	4.7	1430	3	US-09-008-172-2
18	115	4.7	1430	4	US-09-210-361-6
19	115	4.7	1430	4	US-09-740-274-6
20	114	4.6	21	2	US-08-870-827-1
21	114	4.6	21	4	US-09-317-179-1
22	112.5	4.5	2314	4	US-09-268-347-49
23	110	4.5	1073	4	US-09-206-942-47
24	110	4.5	1079	4	US-09-206-942-47
25	108	4.4	834	1	US-08-471-043-21
26	108	4.4	834	1	US-08-471-044-21
27	108	4.4	834	2	US-08-463-483A-21

28	108	4.4	834	2	US-08-471-046A-21	Sequence 21, Appl
29	108	4.4	834	2	US-08-470-566B-21	Sequence 21, Appl
30	108	4.4	834	2	US-08-469-334-21	Sequence 21, Appl
31	108	4.4	834	3	US-09-300-529-21	Sequence 21, Appl
32	108	4.4	1382	3	US-09-057-570-2	Sequence 2, Appl
33	108	4.4	1577	2	US-08-793-824-2	Sequence 2, Appl
34	108	4.4	2123	4	US-08-968-685A-10	Sequence 10, Appl
35	107.5	4.4	725	4	US-09-668-113A-8	Sequence 8, Appl
36	107.5	4.4	2048	4	US-09-268-347-48	Sequence 48, Appl
37	107	4.4	1833	4	US-08-621-944A-4	Sequence 4, Appl
38	107	4.4	1833	4	US-08-945-567D-4	Sequence 4, Appl
39	107	4.4	1992	4	US-08-621-944A-3	Sequence 3, Appl
40	107	4.4	1992	4	US-08-945-567D-3	Sequence 3, Appl
41	105.5	4.3	1541	4	US-08-296-791-3	Sequence 3, Appl
42	105.5	4.3	1541	5	PCT-US95-10661A-3	Sequence 3, Appl
43	105	4.3	1657	3	US-09-057-570-2	Sequence 2, Appl
44	105	4.3	1805	3	US-09-057-570-7	Sequence 7, Appl
45	104	4.2	1507	6	5268270-2	Patent No. 5268270

ALIGNMENTS

```
RESULT 1
US-08-870-827-3
; Sequence 3, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: Tsusaki et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; NUMBER OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSER: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI=2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 455 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-870-827-3

Query Match      100.0%: Score 2454; DB 2; Length 455;
Best local Similarity 100.0%: Pred No. 1.5e+208;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 MNSGDYKEDYGFHAITRADMLKIPGOONSPOFVFQFNASAKINDSADKCYDRSGNIDL 60
|||||
DB 1 MNSGDYKEDYGFHAITRADMLKIPGOONSPOFVFQFNASAKINDSADKCYDRSGNIDL 60

THIS PAGE BLANK (USPTO)

QY 61 DWDSMPLONADGTAAHYHIVSALAGDPKNSDDTPLHLFYOKYGDTSIDSKNAGRY 120
DB 61 DWDSMPLONADGTAAHYHIVSALAGDPKNSDDTPLHLFYOKYGDTSIDSKNAGRY 120
QY 121 FEDMDKFEVNDPYLKYQTOEWSGSATLTKDQVRLFTYDYSNPEDEGTAGAGNOIISTAQ 180
DB 121 FEDMDKFEVNDPYLKYQTOEWSGSATLTKDQVRLFTYDYSNPEDEGTAGAGNOIISTAQ 180
QY 181 VNLSPDPAATLKYDGVSDHKSVPDGGDGTYYONIQQFIDEGKWTISGDNHTLRDPHYVEDK 240
DB 181 VNLSPDPAATLKYDGVSDHKSVPDGGDGTYYONIQQFIDEGKWTISGDNHTLRDPHYVEDK 240
QY 241 GHKYLVEFANTGTDDGYOGDQSFNNKAYYGSDVFFQNEKNKLLQSPKKQIASLANGALG 300
DB 241 GHKYLVEFANTGTDDGYOGDQSFNNKAYYGSDVFFQNEKNKLLQSPKKQIASLANGALG 300
QY 301 IVELADDTYVKSVMPLVASNTVADEVERANIFKMNKKWYLFIDSRSKMTSDGINDKDY 360
DB 301 IVELADDTYVKSVMPLVASNTVADEVERANIFKMNKKWYLFIDSRSKMTSDGINDKDY 360
QY 361 YMLGGGSLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHPGNNVLTSTYMTNRG 420
DB 361 YMLGGGSLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHPGNNVLTSTYMTNRG 420
QY 421 FYPEHSHLRDKLGVNKGSDTSGGENSSGGOFP 455
DB 421 FYPEHSHLRDKLGVNKGSDTSGGENSSGGOFP 455

RESULT 2
US-09-317-179-3

Sequence 3, Application US/09317179
Patent No. 6383769

GENERAL INFORMATION:

APPLICANT: Tsusaki et al.

TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE ACTIVITY

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/317,179

FILING DATE: 24-May-1999

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/870,827

FILING DATE: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TSUSAKI-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 455 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

317-179-3

Query Match 100.0%; Score 2454; DB 4; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.5e-208;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSGDKREDYGFHITRADLKIIPGOONSQFVPPQNASAKINISAKYDYSGNLIDL 60
DB 1 MNSGDKREDYGFHITRADLKIIPGOONSQFVPPQNASAKINISAKYDYSGNLIDL 60
QY 61 DWDSMPLONADGTAAHYHIVSALAGDPKNSDDTPLHLFYOKYGDTSIDSKNAGRY 120
DB 61 DWDSMPLONADGTAAHYHIVSALAGDPKNSDDTPLHLFYOKYGDTSIDSKNAGRY 120
QY 121 FEDMDKFEVNDPYLKYQTOEWSGSATLTKDQVRLFTYDYSNPEDEGTAGAGNOIISTAQ 180
DB 121 FEDMDKFEVNDPYLKYQTOEWSGSATLTKDQVRLFTYDYSNPEDEGTAGAGNOIISTAQ 180
QY 181 VNLSPDPAATLKYDGVSDHKSVPDGGDGTYYONIQQFIDEGKWTISGDNHTLRDPHYVEDK 240
DB 181 VNLSPDPAATLKYDGVSDHKSVPDGGDGTYYONIQQFIDEGKWTISGDNHTLRDPHYVEDK 240
QY 241 GHKYLVEFANTGTDDGYOGDQSFNNKAYYGSDVFFQNEKNKLLQSPKKQIASLANGALG 300
DB 241 GHKYLVEFANTGTDDGYOGDQSFNNKAYYGSDVFFQNEKNKLLQSPKKQIASLANGALG 300
QY 301 IVELADDTYVKSVMPLVASNTVADEVERANIFKMNKKWYLFIDSRSKMTSDGINDKDY 360
DB 301 IVELADDTYVKSVMPLVASNTVADEVERANIFKMNKKWYLFIDSRSKMTSDGINDKDY 360
QY 361 YMLGGGSLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHPGNNVLTSTYMTNRG 420
DB 361 YMLGGGSLNGPHNPINETGLVLMNLDPADLTHYSHCGIPHPGNNVLTSTYMTNRG 420
QY 421 FYPEHSHLRDKLGVNKGSDTSGGENSSGGOFP 455
DB 421 FYPEHSHLRDKLGVNKGSDTSGGENSSGGOFP 455

RESULT 3
US-08-362-232-2

Sequence 2, Application US/08362232
Patent No. 5641667

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Fructosyltransferase Enzyme, Method

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann & Baron

STREET: 350 Jericho Turnpike

CITY: Jericho

STATE: New York

COUNTRY: United States of America

ZIP: 11758

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 MB

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Wordperfect 6.0 for DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/362,232

FILING DATE: 22-December-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CG 125/93

FILING DATE: 23-December-1993

ATTORNEY/AGENT INFORMATION:

NAME: Baron, Ronald J.

REGISTRATION NUMBER: 29,281

REFERENCE/DOCKET NUMBER: 294-29

TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 822-3550

TELEFAX: (516) 822-3582

INFORMATION FOR SEQ. ID NO: 2:

THIS PAGE BLANK (USPTO)

SEQUENCE CHARACTERISTICS:
LENGTH: 543 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
US-08-362-232-2

Query Match 11.5%; Score 281.5; DB 1; Length 543;
Best Local Similarity 26.4%; Pred. No. 1.4e-16;
Matches 120; Conservative 59; Mismatches 184; Indels 91; Gaps 26;

QY 5 DYKEDYGFNHTRADMLKI-----PGQNSP-QFKYPOFNAS-AIKNIDSAGYDK 53
DB 38 DPQSDP-TARWTRDALQIKASDATVAAGNSLPALQTMNPADFPVFN----- 87
QY 54 SGNLIDLVWDSMPLOADGTAAHYGHYVSALADPK---NSDDTPLH-----LFYQKV 106
DB 88 ----PDVWVMDTWLIDKHADQFSTNGWEVIFCLTADPNAGYGFDDRHVARIGFFYRRA 143
QY 107 GDTSI-----DSMNAGRVFED--MDKFVPNDPYLKQTOEMSGSATLTK--DGOVRLFY 157
DB 144 GIRASRRPVNGWGTGCHLFRDQASAGVYAGQYTT--NQAEWSSSLMLQIHGNTVSFVY 201
QY 158 TDVSGNPDGDTGAGN---QIISTAOVNLSDPAATLKVDGVSDBKSVFDGGDGTYYON 213
DB 202 TVAFNRD---ANANNITPPQALITQTLGRIHADFNHVMFTGFAHPLQ-PDGVLYON 257
QY 214 IQQFIDEGKMGISGDNHTLDPHYVEDKGH---KYLVEANT-----GTTDQGGQDSFNK 266
DB 238 GAO-----NEFNRPDFTFEDPKHGVNVMVEGTAGQGVANCTEADLGR-- 306
QY 267 AYYGSDVFEONEKN-KLQSPKKQIASLANGALGIVELADDTYK--SVMKPLVASNTV 323
DB 307 -----PNDNATLQEVLDSCAYYOKANIGLA-IATDSTLSKMKFLSPISANCY 355
QY 324 ADEVERANIFKNNKMYLFTDSRGSKMTSDGINDKD-VYMLGPGGDSLNGPHNPIN-ETG 381
DB 356 NQOTERPQVYLHNKGYIFTISHTTFAA-GVDGPDGVY--GVGDGIRSDFGPMNYGSG 412
QY 382 LVLMNMLPADLHTYSHCGIPHEGNNVLTYSY 415
DB 413 LTWG---NPTDLNTAGTDFDPSPDQNPRAFQSY 443

RESULT 4
US-08-814-196-2
Sequence 2, Application US/08814196
Patent No. 5731173
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Fructosyltransferase Enzyme, Method
TITLE OF INVENTIONS: For its production and DNA encoding the Enzyme.
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann & Baron
STREET: 350 Jericho Turnpike
CITY: Jericho
STATE: New York
COUNTRY: United States of America
ZIP: 11758
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk, 3.5 inch, 1.44 MB
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 6.0 for DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/814,196
FILING DATE: 10-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,232

FILING DATE: 22-December-1994
APPLICATION NUMBER: CU 125/93
FILING DATE: 23-December-1993
ATTORNEY/AGENT INFORMATION:
NAME: Baron, Ronald J.
REGISTRATION NUMBER: 29,281
REFERENCE/DOCKET NUMBER: 294-29
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 822-3550
TELEFAX: (516) 822-3582
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 543 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
US-08-814-196-2

Query Match 11.5%; Score 281.5; DB 1; Length 543;
Best Local Similarity 26.4%; Pred. No. 1.4e-16;
Matches 120; Conservative 59; Mismatches 184; Indels 91; Gaps 26;

QY 5 DYKEDYGFNHTRADMLKI-----PGQNSP-QFKYPOFNAS-AIKNIDSAGYDK 53
DB 38 DPQSDP-TARWTRDALQIKASDATVAAGNSLPALQTMNPADFPVFN----- 87
QY 54 SGNLIDLVWDSMPLOADGTAAHYGHYVSALADPK---NSDDTPLH-----LFYQKV 106
DB 88 ----PDVWVMDTWLIDKHADQFSTNGWEVIFCLTADPNAGYGFDDRHVARIGFFYRRA 143
QY 107 GDTSI-----DSMNAGRVFED--MDKFVPNDPYLKQTOEMSGSATLTK--DGOVRLFY 157
DB 144 GIRASRRPVNGWGTGCHLFRDQASAGVYAGQYTT--NQAEWSSSLMLQIHGNTVSFVY 201
QY 158 TDVSGNPDGDTGAGN---QIISTAOVNLSDPAATLKVDGVSDBKSVFDGGDGTYYON 213
DB 202 TVAFNRD---ANANNITPPQALITQTLGRIHADFNHVMFTGFAHPLQ-PDGVLYON 257
QY 214 IQQFIDEGKMGISGDNHTLDPHYVEDKGH---KYLVEANT-----GTTDQGGQDSFNK 266
DB 258 GAO-----NEFNRPDFTFEDPKHGVNVMVEGTAGQGVANCTEADLGR-- 306
QY 267 AYYGSDVFEONEKN-KLQSPKKQIASLANGALGIVELADDTYK--SVMKPLVASNTV 323
DB 307 -----PNDNATLQEVLDSCAYYOKANIGLA-IATDSTLSKMKFLSPISANCY 355
QY 324 ADEVERANIFKNNKMYLFTDSRGSKMTSDGINDKD-VYMLGPGGDSLNGPHNPIN-ETG 381
DB 356 NQOTERPQVYLHNKGYIFTISHTTFAA-GVDGPDGVY--GVGDGIRSDFGPMNYGSG 412
QY 382 LVLMNMLPADLHTYSHCGIPHEGNNVLTYSY 415
DB 413 LTWG---NPTDLNTAGTDFDPSPDQNPRAFQSY 443

RESULT 5
US-09-503-172A-2
Sequence 2, Application US/09503172A
Patent No. 6284510
GENERAL INFORMATION:
APPLICANT: ITO, Tetsuya
APPLICANT: FUJITA, Koki
APPLICANT: HARA, Kozo
APPLICANT: TONOUKA, Takashi
APPLICANT: SAKANO, Yoshiyuki
TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE GENE
FILE REFERENCE: 10749-0001-0
CURRENT APPLICATION NUMBER: US/09/503,172A
CURRENT FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: JP 160416/1999
PRIOR FILING DATE: 1999-06-08

THIS PAGE BLANK (USPTO)

NUMBER OF SEQ ID NOS: 9
 SOFTWARE: Patentin Ver. 2.1
 SEQ ID NO: 2
 LENGTH: 578
 TYPE: PRT
 ORGANISM: Arthrobacter sp.
 US-09-503-172A-2

Query Match 11.3%; Score 278.5; DB 4; Length 578;
 Best Local Similarity 23.7%; Pred. No. 2.7e-16;
 Matches 125; Conservative 67; Mismatches 193; Indels 143; Gaps 25;

8 EDYGAHTTRADMLK-----GQOQNSPOFK-----VPOFNASAIKNIDSAKGYDSS 54
 57 EDDFAKTRADARLOKMSDPTAPSRKSMSPASVTMPVPO----- 98
 55 GNLIDL-----VMSWPLQNDGTRANYHGYHSALAGDKP-NSDD-----TPHLFYQ 104
 99 -DFPMSNEQWVWMDTFLTDADANQSVNGWEIIFSLVADRNLTGDDRHVFAKIGFYR 157
 105 KYGDSID-----SKKNAGRVT-EDMDKRVPRNDPIKYOTQWSSGATLTGQVRLFT 158
 158 PAGVPAERPENGGVTYGLVKEGVTGQIFEDQSFHQTO-WSGSARVSKNGEIKLFT 216
 159 DYSGNPEDGTCG-----GNQISTAQVNLSPDAATLKVDGSDHKSVPDGGDGYONI 214
 217 DVAFRNSDGTNIRKIDPRIALSVKRVANKKGVLLTGKNTDLDQ-----ADGITYOTG 272
 215 QQFIDEGKMWISGDNHTLRDPHYVEDKGH--KYLVFEAN-----TGTTD--GY-QGD 260
 273 AQ-----NEFENFRDPFTFEDPAHGEFTFVVEGNSAMQREIATCNEDLGRQGD 323
 261 QSFNNKAYGSGDVEFQNEKKNLLOSPKKQIASLANGALGYELADDTYKSVMKPLVAS 320
 324 PYAETVDVNASGATYQIGNGLAKAKNKOL-----TEWEFLPILISA 366
 321 NTVADEVERANIFKNNKMYLFTDSRGSKMTSDGINDKDYMLPGGSDLNPHNIN-E 379
 367 NCVIDQTRPOLYFEDGKSYLFTISHRGTFAGLDGPEGV--GPGGDSIRSDYQPLNG 424
 360 TGLVLMNLDPADLT-----HTYSHGCIHPPEGNNVLTSYMTNRGF 421
 425 SGLALG---NPTNLNLFGLGQFPAPDFNQHPGHQAYSHVMP-----GGLVQSFIDTIG- 475
 422 YPEHSHLRD-----KLGVNIKSDIS-----GGENSSGQCGFP 455
 476 --THDFVRGGTLAPTVKMDIGVGDPTKTAVDYSYSGELGGMADIP 521

RESULT 6
 US-08-381-936-2
 Sequence 2, Application US/08381936
 Patent No. 5792923
 GENERAL INFORMATION:
 APPLICANT: ROBER, Manuela
 APPLICANT: GEIER, Gebhardt
 APPLICANT: GEIDER, Klaus
 APPLICANT: WILMITZER, Lothar
 TITLE OF INVENTION: DNA sequences which lead to the
 TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing
 TITLE OF INVENTION: these sequences as well as a process for preparing
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ostroienk, Faber, Gerb & Sofien
 STREET: 1180 Avenue of the Americas
 CITY: New York
 STATE: NY
 COUNTRY: US
 ZIP: 10036-8403

COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/381,936
 FILING DATE: 09-FEB-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: EP 93 02110
 FILING DATE: 09-AUG-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: DE P42270618
 FILING DATE: 08-DEC-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Meliman, Edward A.
 REGISTRATION NUMBER: 24,735
 REFERENCE/DOCKET NUMBER: P/951-108
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 382-0700
 TELEFAX: (212) 382-0888
 TELEX: 236925
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 415 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-381-936-2

Query Match 10.4%; Score 255; DB 1; Length 415;
 Best Local Similarity 25.4%; Pred. No. 2e-14;
 Matches 115; Conservative 64; Mismatches 141; Indels 132; Gaps 29;

9 DYGFHAI--TRADMLKIPGOQNSPOFKVPOFNASAIKNIDSAKGYDCKGNLIDL----- 61
 3 DYVKKFTLTRADALKV--HEDDPTTTP-----VIDIAPVMS 39
 62 ---VMSWPLQNDGTRANYHGYHSALAGDKPNSDDTPHL----- 100
 40 EEVFTIDTMLRPFDEIISVNGWCITFTLTAD-RNTDNGQFODENGNNYDTRDWEHRIG 98
 101 ----LFYOKVGDTSIDSKNAGRVFEDMDKFPVNPPLYETKYOTQWSSGATLTGK-DQVR 154
 99 RARICWYSRTG---KDMIFGSRV--AEGVAPF-----TRENAGITLLNDRDID 145
 155 LEYTDYSGNPEDGTCG-----GNQISTAQVNLSPDAATLKVDGSDHKSVPDGGDGYV 211
 146 LYTTCVT---PGATIAKVRGKIIVTSDQSVSL-----BGFOQVTSLF-SADGTIY 190
 212 QNIQCFIDEGKMTSGNHTLRDPHYVEDK--GHKYLVFEANTGTTCGYQGDQSFNNKAYY 269
 191 QTEQ---NAFW-----NFRDPSPIIDRDGKLYMIEGNVA--GPRSSHET-TOAEM 237
 270 GGSDFVONEKKNLLOSPKKQIASLANGALGYELADDTYK--SVKKPLVASNTVADEV 327
 238 GNPFCYEDVGAKKYA-----GCVGLA-VAKDLSGSEMOLLPLITAVGYVNDOT 286
 328 ERANIFKNNKMYLFTDSRGSKMT-SDGINDKD-VYMLGPGGSLNGPHNINETGLYL- 384
 287 ERHEHFEQGGKYLLFTISH--KYTFADNLTGPGGV--GVSQKLGLPYTPMSSGLVIG 342
 385 NMNLDPADLTHYSHGCIHPPEGNNVLTYSYM 416
 343 NPSSQP---FOTYSHVMP-----NGLVTSFI 366

RESULT 7
 US-08-943-374-2
 Sequence 2, Application US/08943374
 Patent No. 6028249
 GENERAL INFORMATION:
 APPLICANT: ROBER, Manuela
 APPLICANT: GEIER, Gebhardt
 APPLICANT: GEIDER, Klaus

THIS PAGE BLANK (USPTO)

APPLICANT: WILMITZER, Lothar
TITLE OF INVENTION: DNA sequences which lead to the
formation of polyfructans (levans), plasmids containing
TITLE OF INVENTION: these sequences as well as a process for preparing
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ostrolenk, Faber, Gerb & Coffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,374
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/381,936
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P42270618
FILING DATE: 08-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Mellman, Edward A.
REGISTRATION NUMBER: 24,735
REFERENCE/DOCKET NUMBER: P/951-108
COMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 415 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-374-2
Query Match 10.4%; Score 255; DB 3; Length 415;
Best Local Similarity 25.4%; Pred. No. 2e-14;
Matches 115; Conservative 64; Mismatches 141; Indels 132; Gaps 29;
QY 9 DYGFANI--TRADMLKIPGOONSPOFKVPOFNASAIKNIDSAGKYDKSGNLLID-- 61
DB 3 DYNKKPTLMTTRADALKV--HEDPTTTP--VIDIAFPVMS 39
QY 62 ---VWDSMPLQADGTAANYHGYHYSALAGDPKNSDTPH-- 100
DB 40 EEFYIWDWPLRFDEIISVNCWCIIFLTAD-RMTDNPQFODENGNYDITRDWEDRHG 98
QY 101 ---LFYKVGPTSIDSMKNAGRVEDMDKFEVNDPYLKYQOEWGSLTKRD-QQVR 134
DB 99 RARICWYSRTG---KDMIFGSRV--AEGVAP--TREAAGTPTLLINDRDID 145
QY 155 LFYTYDSGNPEDGGTGA---GNOIISTAVNLSPDPAATLVGVSDHKVFEPGGGTGY 211
DB 146 LYTCVT---PGATIAKVRGKIIVTSDQSVSL---EGFOQVTSLF-SADGTIY 190
QY 212 QNIOGFIDGKWIISGDNHTLRDPHYEDK--GHKYVFEANNGTDTGOGDOSFNKKAY 269
DB 191 QTEEQ---NAFW---NFRDPSPEIDNNDGKLYMLFEGNVA---GPRGSHET-TQAEK 237
QY 270 GSDVEFQNEKKNKLLQSPKKQIASLANGALGIVEIADDDTVK--SYMKPLVASNTVADVEY 327
DB 238 GNVPGYEVGCAKQYA---GGVGLA-VAKDLSGSEWQILPLPLTAIVGVNDQT 286
QY 328 ERANIFKNNKWLFTDSRGSKMT--SDGINDKD-VYMLPGGDSLNGPHNPINETGLV- 384

DB 287 ERPHFVQDGXYVLTFTISH--KYTFADNLTPDPGV--GFVSDMLTGTYTMSNGVLG 342
QY 385 NMNIDPADLFTYTHCGCPHPEGNNVLTSM 416
DB 343 NPSSQP---FQYISHYVP-----NGLVTSPI 366
RESULT 8
US-09-007-999-2
Sequence 2, Application US/09007999
Patent No. 6087559
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OF INVENTION: Substitutes for Modified Starch and
TITLE OF INVENTION: Latexes in Paper Manufacture
FILE REFERENCE: 0356D
CURRENT APPLICATION NUMBER: US/09/007,999
CURRENT FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/478,704
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1475
TYPE: PR1
ORGANISM: Streptococcus mutans
US-09-007-999-2
Query Match 5.2%; Score 128; DB 3; Length 1475;
Best Local Similarity 20.4%; Pred. No. 0.022;
Matches 79; Conservative 48; Mismatches 102; Indels 158; Gaps 22;
QY 68 LONADGTANYH-----YHVSALADDPKNSDTPHLPYQKYGDSISWKNA 117
DB 1131 LKNEDGTYA-YGNDGRRENGYQFMGWRHFNNGE-----MSVGLTVID----- 1176
QY 118 GRV--FEDMDKFPNDPYLKYQTOEMSGSATLTKDGOVRLFTDYSGN----- 163
DB 1177 GOVYFEDMG-----YQAK--GKFTYTAGKIR-YFDKSGNNYRRTIENEG 1222
QY 164 ---PEGGTGAGAGQIISTQVNLSPDPAATLVKDYVSDHK---SVFDDGGTYQNIQ 215
DB 1223 KWLVEGGAANTVSSQITN-QQHLFYRANGVQVKEFTDHHGRISYDNGSDQIRN-- 1279
QY 216 QFI--DEGKWTSGDNHTLRDPHYEDKH-----KLVFEAN----- 250
DB 1280 RFVNNAGQWF-----YFDNNGYAVTGARTINGQLLYFRANGVQKGEFVTDRY 1328
QY 251 --TGTTDGOYQDQ-----SFNNKAY-----YGGSDVEFQNEKKNKLLQSP 287
DB 1329 GRISYDNGSDQIRNFRVRAAGQWFYFDNNGYAVTGARTINGQHLTYR----- 1378
QY 288 KQQLASLANGALGIVEIADDTYKSVKPLVASNTVADVERANIFKNNKWLFTDSRG 347
DB 1379 ---ANG---VQYKGEFVTDHGRIRISYDNGSDQIRNFRVRAAGQWFYF-DNNG 1426
QY 348 SKMTSDGINDKDVYMLPGGDSLNGPH 374
DB 1427 YAVT-----GARTINGOH 1439
RESULT 9
US-09-210-361-2
Sequence 2, Application US/09210361
Patent No. 6284479
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OF INVENTION: Substitutes for Modified Starches and
TITLE OF INVENTION: Latexes in Paper Manufacture
FILE REFERENCE: 0357CR
CURRENT APPLICATION NUMBER: US/09/210,361
CURRENT FILING DATE: 1998-12-11

THIS PAGE BLANK (USPTO)

Query Match	5.28;	Score 128;	DB 4;	Length 1475;
Best Local Similarity	20.48;	Pred. No. 0.022;		
Matches 79;	Conservative 48;	Mismatches 102;	Indels 158;	Gaps 22

RESULT 10
US-09-740-274-2
Sequence 2, Application US-09-740-274
Patent No. 6465203
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OR INVENTION: Glucan-containing Compositions and Paper
FILE REFERENCE: 0357CRD
CURRENT APPLICATION NUMBER: US/09/740,274
CURRENT FILING DATE: 2000-12-19
PRIOR APPLICATION NUMBER: 09/220,361
PRIOR FILING DATE: 1998-12-11
PRIOR APPLICATION NUMBER: 09/007,999
PRIOR FILING DATE: 1998-01-16
PRIOR APPLICATION NUMBER: 08/478,704
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: 09/009,620
PRIOR FILING DATE: 1998-01-20
PRIOR APPLICATION NUMBER: 08/485,243
PRIOR FILING DATE: 1995-06-07

Query Match	5.28;	Score 128;	DB 4;	Length 1475;
Best Local Similarity	20.48;	Pred. No. 0.022;		
Matches	79;	Conservative	48;	Mismatches 102;
				Indels 158;
				Gaps 22

RESULT 11
 US-08-296-791-6
 Sequence 6, Application US/08296791
 Patent No. 624537
 GENERAL INFORMATION:
 APPLICANT: St. Geme III, Joseph W.
 APPLICANT: Falkow, Stanley
 TITLE OF INVENTION: Haemophilus Adherence and Penetration
 TITLE OF INVENTION: Protein
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 STREET: 4, Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/296,791
 FILING DATE: 25-AUG-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Treacattin, Richard F.

THIS PAGE BLANK (USPTO)

```

;
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELE: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
;
US-08-296-791-6

Query Match          4.7%; Score 116.5; DB 4; Length 1848;
Best Local Similarity 20.8%; Pred. No. 0.32;
Matches 105; Conservative 64; Mismatches 181; Indels 155; Gaps 28;

QY 25 GQONSPOKVPQFNAS--AIKNIDSAGYDKSGNLIDLDVWDSWPLQANADGTAAANYHGYH 82
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 GDSGSPLEFYDREKGMWFLGSDYDFWAGYNNKS-----WQEMNIYKHEFAEKIYQOYS 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 83 IVSAL-----AGDPKNSD-----DTPLH----- 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 AGSLIGSNQYTWQATGSTTTGGGEPLSVLDLTDGKDPNPKSITLKGSGTLTLNNHI 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 101 -----LF-----YQKVDTSIDSMKNAQRFEDMDKFPV--NDPYLKYQTQEMSGSATL 147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 DQAGAGLFEEDGYEVKGTSDSTTWKAGAVSAD-GKTVWKVHNP--KYDRLAKIGKGTL 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 148 TKDGOVRLFTYDTSNPE-----DG-----GTGAGNOIISTAQVNL-SQDDAATLKYD 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 VVEGK-----GKNGLLKVGDGYILKOKADANNKVOAFSGVIGSGSTLVLLND 516
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 195 GVSDDKSVFDG-----GDGTYQNTQGFIDEGKMTSGDNHRLRDPHYEDGKHYL 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 KOYDPSNITFGRGRLDNGSLTFDHLR-IDDGARV--NHNMTNSNITIGES-L 572
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 246 VFEANTGTIDGYQGDOS-----FN--NKAYYGSDFVFEONEKNKLQSPK 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 573 ITNPNTITSYNIEADDDHPLRISIPYQLYFNQDNRSY-----TLKGASTRSEL 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 289 KOIASLANGALGIVELADDTYKSVKPLVASNTVADEVERANIFKMNKWLFTDSRGS 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 PONGSESNEMWLYMGRTSDAARNVNNH-----NNERNNGF--NGYGEETKAT 674
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 349 K-----MTSDGINDKDVYMLGPGGSLNGPHNPINETGLVLMNIDPADLFTYSHCGI- 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 QNGKLVTFNGKSDONRFL-TGGTNLNGDLN-VEKGTFLISGRPTP-----HARDIAGIS 728
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 403 -----PH-PEGNNVYLTSYMTNRGF 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 729 STKDPHFTENNEVVEEDWIMNRNF 753
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
PCT-US95-10661A-6
; Sequence 6, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: Flehr, Hohbach, Test, Albrighton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER: IBM PC compatible
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Tregetan, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELE: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
;
PCT-US95-10661A-6

Query Match          4.7%; Score 116.5; DB 5; Length 1848;
Best Local Similarity 20.8%; Pred. No. 0.32;
Matches 105; Conservative 64; Mismatches 181; Indels 155; Gaps 28;

QY 25 GQONSPOKVPQFNAS--AIKNIDSAGYDKSGNLIDLDVWDSWPLQANADGTAAANYHGYH 82
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 GDSGSPLEFYDREKGMWFLGSDYDFWAGYNNKS-----WQEMNIYKHEFAEKIYQOYS 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 83 IVSAL-----AGDPKNSD-----DTPLH----- 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 AGSLIGSNQYTWQATGSTTTGGGEPLSVLDLTDGKDPNPKSITLKGSGTLTLNNHI 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 101 -----LF-----YQKVDTSIDSMKNAQRFEDMDKFPV--NDPYLKYQTQEMSGSATL 147
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 DQAGAGLFEEDGYEVKGTSDSTTWKAGAVSAD-GKTVWKVHNP--KYDRLAKIGKGTL 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 148 TKDGOVRLFTYDTSNPE-----DG-----GTGAGNOIISTAQVNL-SQDDAATLKYD 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 VVEGK-----GKNGLLKVGDGYILKOKADANNKVOAFSGVIGSGSTLVLLND 516
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 195 GVSDDKSVFDG-----GDGTYQNTQGFIDEGKMTSGDNHRLRDPHYEDGKHYL 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 KOYDPSNITFGRGRLDNGSLTFDHLR-IDDGARV--NHNMTNSNITIGES-L 572
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 246 VFEANTGTIDGYQGDOS-----FN--NKAYYGSDFVFEONEKNKLQSPK 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 573 ITNPNTITSYNIEADDDHPLRISIPYQLYFNQDNRSY-----TLKGASTRSEL 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 289 KOIASLANGALGIVELADDTYKSVKPLVASNTVADEVERANIFKMNKWLFTDSRGS 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 PONGSESNEMWLYMGRTSDAARNVNNH-----NNERNNGF--NGYGEETKAT 674
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 349 K-----MTSDGINDKDVYMLGPGGSLNGPHNPINETGLVLMNIDPADLFTYSHCGI- 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 QNGKLVTFNGKSDONRFL-TGGTNLNGDLN-VEKGTFLISGRPTP-----HARDIAGIS 728
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 403 -----PH-PEGNNVYLTSYMTNRGF 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 729 STKDPHFTENNEVVEEDWIMNRNF 753
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-09-134-001C-4469
; Sequence 4469, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
```

CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4469
LENGTH: 513
TYPE: PRF
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4469

Query Match 4.7%; Score 116; DB 4; Length 513;
Best Local Similarity 20.2%; Pred. No. 0.051;
Matches 96; Conservative 62; Mismatches 144; Indels 174; Gaps 22;

QY 86 ALADPKNSDDTPLHLEFYQKVGDTSIDSMKNAAGVFEEDMDKFPNDPYLKYQTEWMSGSA 145
D 57 ALKELPK-SENKKHKKYDKVYDTEKD-----NKGFHTTLPKVG- 97
QY 146 TLTDGQVRLFTYYSNP--EDGCTGAGNOIISTAOYNLSDPAA----- 189
D 98 TYADPKVKV-HTKKESKVLVNGDTDA-KKVQPTNKVSIKESATDKAFAIKIDROKA 155
QY 190 -TLKVDGVSDFKSYFDGDDGVYQNIQ-----OFIDE----- 220
D 156 KNLKSDYIKTKKVELDEKKNYVNIETITSPKISHNNKVIDAETGOVVKLNMKIRAA 215
QY 221 ---GKWISGDN-----TLRDPHYEDKGKLYVEANTGTDD 256
D 216 TTGKGVLGDTKQININSVSGYALODLTQCGTSLAANYDANTGOAVLMDKDNFDD 275
QY 257 YQ-----GDSTNNKAYYGGSVF-----FQNEKNLLQSPK 288
D 276 EORAGVDANYAKETDYDKNTFFGRESTYDNO---GSPILSLAHVNNQGDNR----- 325
QY 289 KOIASLANGA--LGIVELADDTYKSVKPLVANSNTVADVEY-----ERANIFKNNKMY 340
D 326 ---NNAMIDDKMIGYDGDGTFALSGANDVVAHEITHTGTOGTALVYRSOSGA 378
QY 341 L--FTDSRSGKM-TSDGINDKDYVMGPGDLSLGNPNINFTGLVLMNLDPADLTHT 396
D 379 LNESFSDVFGYVDEDFLGMEDVYTPGVGDALRSMSNP--ERFGQPSHMN---DEYVT 433
QY 397 ---YSHCGIPHEGNNVY-----LTSYMTNGEYFEHSHLR 430
D 434 NSDNGVHTNSGIRKKAAYNTIRSIGKORSEQIYRALTLYLTSNSDFODAKASIQ 489

RESULT 14
US-08-296-791-4
Sequence 4, Application US/08296791
Patent No. 6245337
GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
APPLICANT: Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trecalet, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1545 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-296-791-4

Query Match 4.7%; Score 115.5; DB 4; Length 1545;
Best Local Similarity 20.8%; Pred. No. 0.3;
Matches 104; Conservative 55; Mismatches 193; Indels 149; Gaps 23;

QY 25 GQNSPOFKYQENAS--AIKNIDSAGYDKSGMLIDLVDSPLOADGTAAHYGYH 82
D 290 GDSSGPLEFYDREKGMFLGSDYMWAGYNNKS-----WQENNIYKPEFAEKTYEYS 342
QY 83 IVSALADGPKNSDDTPLHLEFYQKVGDTSIDSMKNAAGVFEEDMDKVF 128
D 343 AGSL-----IGSKTDVSWSSNGKSTITGGEKSLANDLAGDKD-- 381
QY 129 PNDPYLKYQTEWMSGATLTKD---GOVRLFTYDYSNPEDGTF-----GAGNOIIST 178
D 382 PN--HGKSVTFEFGSLTLNNDGAGGLE---FEGDYEVKIGSDNTMTWAGVSAEG 436
QY 179 AQV-----NLSDPAAATLKVDSVSKSVFDGDDGVYQNIQOFIDEKMW----- 223
D 437 KTVWVKVNPQYDLRIAKIGKGTLLIVEGTGDKGSLKVGDDGVILK--OOTNGSGHAFASV 495
QY 224 --ISG-DNHTLRDPHYVE-----DKGHLYLVEANTG 252
D 496 GIVSGRSLVLDNDKQVDPNISYFGFRGRGLDLNGSLTFPHIRINIDEGALVNHSHSKH 555
QY 253 TTGQYGDQSF--NNKAYV-----GSDVFEQNEKNLLQSPKQIA 292
D 556 STVITGDNLTIDPNNWSIYYVKPLEDDNPYAIRQIKGYQLYF--NEENFTYALKKQAS 614
QY 293 SLANGALGIVELADDTYKSVKPLVANSNTVADVEYERANIFKNNKMYLFTDSRGS----- 348
D 615 IRSEFPQNRGESSNWSMLYGTAKDAQNA---MNHINERMGFPNGYFGEEGKNGN 670
QY 349 KMTSDGINDKDYVMGPGDLSLGNPNINFTGLVLMNLDPADLTHYSHCGIP----- 403
D 671 LNVTFKKSSEGNREL-TGNTNLNDLN-VQCGTLFLSGRTP---HARDIAGISSTKK 724
QY 404 --HPEGNNVLTSMNNGF 421
D 725 DSHFSENNEVVEEDWIRNF 745

RESULT 15
PCT-US95-10661A-4
Sequence 4, Application PC/TUS9510661A
GENERAL INFORMATION:
APPLICANT: Washington University, et al.
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 16-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: PP-59941/RET
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1545 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-4

Query Match

4.7%; Score 115.5; DB 5; Length 1545;

Best Local Similarity 20.88; Pred. No. 0.3;
Matches 104; Conservative 55; Mismatches 193; Indels 149; Gaps 23;

QY 25 GQONSPOFKYPOFVNAS--AIKNIDSAGYDKSGNLIDLDVWDSWPLQADGTANYHGYH 82
DB 290 GDGSGPLFVYDREKGNLFLGSIYMAGYNKS-----WQEWNIYKPEFAEKIYECYS 342
QY 83 IVSALAGDPKNSDDTPLHFLFYQKYVDTSIDSMKNAGR-----VFEDMDKFV 128
DB 343 AGSL-----IGSKTDYSWMSNGKSTITGGEKSLNVDLADGDKD-- 381
QY 129 PNDPYLKYQYQEWGSGATLTKD---GQYRLFYTDYSGNPEDEGT-----GAGNQIIST 178
DB 382 PN--HGKSVTFEGSGTLTLNNNIDGAGGLE--FEGDYEVKGTSDNTYMKAGVSYAEG 436
QY 179 AQV-----NLSQPDATLKVDSVDHKSVEFGDGGTVYQNIQOFTDEGKW----- 223
DB 437 KTYVMKYNPOYDLAKIGKGLIVEGTGDNKGS�KYGDGYILK--QQTNGSGQHAFASV 495
QY 224 --ISG-DNHTLRDPHYE-----DKGHRKYLVEEANTG 252
DB 496 GIVSGRSTLYLNDKQYDPNSIYFGRGRLDNGNSLTFDHIRNIDEGARLVNHSTSKH 555
QY 253 TTDDYQGDQSF---NNKAY-----GSDVFFQNEKNLLQSPKKQIA 292
DB 556 STVYITDNLITDPNNVSIYVKPLEDDNPYAIRQIKYQOLYF--NEENPTYVALKRDAS 614
QY 293 SLANGALGIVELADYVKSVMKPLVASNTVADEVERANIFKNNKMYLFTDSRGS--- 348
DB 615 IRSFEFPQNRGESSNSMLYMGTEKADAKNA---MNHINERMRNGFNGYEEEGKNNGN 670
QY 349 -KMTSDGINDKDYVMLGPGDLSLNGPHNPINETGLVLMNMLDPADLTHYSHCGIP--- 403
DB 671 LNVTFKCKSEQNRFLL--TGGTNLNGDLN--VOQGTFLFSGRPTP---HARDIAGISSTKK 724
QY 404 ---HPEGNNVYVLSYMTNRGF 421
DB 725 DSHFSENNEVYVEDDWINRNF 745

Search completed: June 16, 2003, 12:11:35
Job time : 44.0246 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:07:33 ; Search time 69.9283 seconds
(without alignments)
695.665 Million cell updates/sec

Title: US-09-986-682B-3
Perfect score: 2454
Sequence: 1 MNSGDYKEDYGFRAHITRADM.....NIKSGDTSGGENSSGQGF 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues
Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2454	100.0	455	10	US-09-986-682B-3
2	649.5	26.5	792	10	US-09-995-587A-11
3	606.5	24.7	789	10	US-09-995-587A-11
4	129.5	5.3	1411	9	US-10-080-505-17
5	128	5.2	1475	10	US-09-740-274-2
6	126	5.1	2076	10	US-09-815-242-5815
7	126	5.1	2186	10	US-09-815-242-5815
8	125	5.1	1981	9	US-09-928-457-38
9	125	5.1	2015	9	US-10-066-551-1
10	120	4.9	773	12	US-10-067-385-8
11	116.5	4.7	1848	9	US-09-839-996-6
12	116.5	4.7	1848	9	US-10-080-505-6
13	115.5	4.7	1545	9	US-09-839-996-4
14	115.5	4.7	1545	9	US-10-080-505-4
15	115	4.7	467	10	US-09-912-020-381
16	115	4.7	1430	10	US-09-740-274-6
17	114.5	4.7	1434	9	US-10-080-505-9
18	114	4.6	21	10	US-09-986-682B-1
19	113	4.6	1325	10	US-09-741-669-304

20	113	4.6	1436	9	US-10-080-505-13	Sequence 13, Appl
21	110.5	4.5	546	10	US-09-815-242-5510	Sequence 5510, Ap
22	110.5	4.5	549	10	US-09-815-242-12592	Sequence 12592, A
23	108	4.4	915	10	US-09-817-514A-6	Sequence 6, Appli
24	108	4.4	978	10	US-09-815-242-5456	Sequence 5456, Ap
25	108	4.4	1001	10	US-09-815-242-12686	Sequence 12686, A
26	108	4.4	2122	9	US-09-813-214A-9	Sequence 9, Appli
27	106.5	4.3	650	10	US-09-801-368-430	Sequence 430, App
28	106.5	4.3	2039	9	US-10-192-584-7	Sequence 7, Appli
29	106	4.3	2516	10	US-09-817-514A-2	Sequence 2, Appli
30	105.5	4.3	1541	9	US-09-839-996-3	Sequence 3, Appli
31	105.5	4.3	1541	9	US-10-080-505-3	Sequence 3, Appli
32	105	4.3	727	10	US-09-815-242-5782	Sequence 5782, Ap
33	105	4.3	727	10	US-09-815-242-12777	Sequence 12777, A
34	105	4.3	1023	9	US-09-884-696-5	Sequence 5, Appli
35	104	4.2	1344	9	US-09-738-626-6888	Sequence 6888, Ap
36	103.5	4.2	1018	10	US-09-815-242-5797	Sequence 5797, Ap
37	103.5	4.2	1018	10	US-09-815-242-12838	Sequence 12838, A
38	103.5	4.2	2710	9	US-10-011-366-6	Sequence 6, Appli
39	103	4.2	550	9	US-10-125-692-22	Sequence 22, Appl
40	102	4.2	999	9	US-09-884-465A-376	Sequence 376, App
41	102	4.2	1391	9	US-10-080-505-11	Sequence 11, Appl
42	102	4.2	1391	9	US-10-080-505-15	Sequence 15, Appl
43	102	4.2	1752	10	US-09-841-132-180	Sequence 180, App
44	102	4.2	6281	10	US-09-815-242-12996	Sequence 12996, A
45	101.5	4.1	1027	1	US-08-781-986A-5254	Sequence 5254, Ap

ALIGNMENTS

RESULT 1
US-09-986-682B-3
Sequence 3, Application US/09986682B
Patent No. US20020115182A1
GENERAL INFORMATION:
APPLICANT: KENKUYO
KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/986,682B
FILING DATE: 06-NO. US20020115182A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO. 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 3
US-09-986-682b-3
Query Match 100.0%; Score 2454; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 2,5e-189;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNSGKEDYGFANHTRADMLKIPGOONSPOKVPQFNASAIKNIDSAKGYKSGNLIDL 60
DB 1 MNSGKEDYGFANHTRADMLKIPGOONSPOKVPQFNASAIKNIDSAKGYKSGNLIDL 60
QY 61 DWSDWSPLOQADGTAAHYGHYIVSALAGDPKNSDDTPLHLFYQKVDTSIDSMKNAGRY 120
DB 61 DWSDWSPLOQADGTAAHYGHYIVSALAGDPKNSDDTPLHLFYQKVDTSIDSMKNAGRY 120
QY 121 FSDMKFVNDPYLKYQOEMSGSATLTFDGOVRLFTYTSQNPEDGGGAGANOIISTAQ 180
DB 121 FSDMKFVNDPYLKYQOEMSGSATLTFDGOVRLFTYTSQNPEDGGGAGANOIISTAQ 180
QY 181 VMLSDPDAATLKYDGVSDHKSVEFDGDDGVYONIOQFIDEKMWISGDNHTLBDPHYVEDK 240
DB 181 VMLSDPDAATLKYDGVSDHKSVEFDGDDGVYONIOQFIDEKMWISGDNHTLBDPHYVEDK 240
QY 241 GHKYLVEFANTGTGCGOQSFNNKAYYGGSDVFEQNEKNKLLQSPKKQIASLANGALG 300
DB 241 GHKYLVEFANTGTGCGOQSFNNKAYYGGSDVFEQNEKNKLLQSPKKQIASLANGALG 300
QY 301 IVELADDDYVKSVMKPLVANSNTVADEVERANIEFKMNNKWLFTDSGSKMTSDGINDKDY 360
DB 301 IVELADDDYVKSVMKPLVANSNTVADEVERANIEFKMNNKWLFTDSGSKMTSDGINDKDY 360
QY 361 YMLGPGGSLNGPHNPINETGLVLMNNLDPADLTHYSHGCIPIHPEGNNVLTSTYTNNG 420
DB 361 YMLGPGGSLNGPHNPINETGLVLMNNLDPADLTHYSHGCIPIHPEGNNVLTSTYTNNG 420
QY 421 FYPEHSHRLDKLVNIKSDTSGGENSSGOGQFP 455
DB 421 FYPEHSHRLDKLVNIKSDTSGGENSSGOGQFP 455
RESULT 2
US-09-995-587A-11
Sequence 11, Application US/09995587A
Patent No. US20020127681A1
GENERAL INFORMATION:
APPLICANT: VAN HIJUM, SACHA ADRIANUS FOKKE TACO
APPLICANT: VAN GEEL-SCHUTTEN, GERITDINA HENDRIKA
APPLICANT: DIJKHUIZEN, LUBBERT
APPLICANT: RAHAOUTI, HAKIM
TITLE OF INVENTION: NOVEL FRUCTOSYLTRANSFERASES
FILE REFERENCE: BO43667-CIP
CURRENT APPLICATION NUMBER: US/09/995,587A
CURRENT FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: 09/604,958
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: EPO 00201872.9
PRIOR FILING DATE: 2000-05-25
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 792
TYPE: PRT
ORGANISM: Lactobacillus reuteri
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (495)..(496)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (737)
OTHER INFORMATION: Thr or Pro
US-09-995-587A-11

Query Match 26.5%; Score 649.5; DB 10; Length 792;
Best Local Similarity 37.8%; Pred. No. 7.2e-44;
Matches 158; Conservative 77; Mismatches 132; Indels 51; Gaps 17;
QY 27 QNSPQFKVPQFNASAIKNIDSAKGYD-KSGNLIDLDVWDSWPLQNA-ADGTAAHYGHYIV 84
DB 27 QNSPQFKVPQFNASAIKNIDSAKGYD-KSGNLIDLDVWDSWPLQNA-ADGTAAHYGHYIV 84
QY 203 KODPYAIPYSNAKIEKKNPATYVDAQTKAAHLDVWDSWPLQNPATYVQVYVSWGXYLV 262
DB 203 KODPYAIPYSNAKIEKKNPATYVDAQTKAAHLDVWDSWPLQNPATYVQVYVSWGXYLV 262
QY 85 SALAGDPKN-SDDTPLHLFYQKVDTSIDSMKNAGRVEFEDMDKFPNDPYLKYQOEMSG 143
DB 85 SALAGDPKN-SDDTPLHLFYQKVDTSIDSMKNAGRVEFEDMDKFPNDPYLKYQOEMSG 143
QY 263 IAMGIIPIPSPTGDNHTIYLLNKGNDPSHMRNAGSIFGTETNV-----FOEMSG 313
DB 263 IAMGIIPIPSPTGDNHTIYLLNKGNDPSHMRNAGSIFGTETNV-----FOEMSG 313
QY 144 SATLTKDGOVRLFTY-----DYSGNPEDGGGAGANOIISTAQVNLSDPDAATLKYDGYSD 198
DB 144 SATLTKDGOVRLFTY-----DYSGNPEDGGGAGANOIISTAQVNLSDPDAATLKYDGYSD 198
QY 314 SAIVVDDGTIQLEFFISNTSDYKLN-----DQLATATLNLNVDDG-VSISKYSDN 363
DB 314 SAIVVDDGTIQLEFFISNTSDYKLN-----DQLATATLNLNVDDG-VSISKYSDN 363
QY 199 HKSVEFDGDTYVYONIOQFIDEKMWISGDNHTLBDPHYVE-DKHKYLVEFANTGTGCGY 257
DB 199 HKSVEFDGDTYVYONIOQFIDEKMWISGDNHTLBDPHYVE-DKHKYLVEFANTGTGCGY 257
QY 364 YQVLFEE-GDGFHYQYEQFAN-GKRENDYCYLRDPHYVQLENGDRYLVFEANTGTED-Y 420
DB 364 YQVLFEE-GDGFHYQYEQFAN-GKRENDYCYLRDPHYVQLENGDRYLVFEANTGTED-Y 420
QY 258 QGDSQFNNKAYYGGSDVFEQNEKNKLLQSPK-KOIASLANGALGIVELADDDY---VKS 313
DB 258 QGDSQFNNKAYYGGSDVFEQNEKNKLLQSPK-KOIASLANGALGIVELADDDY---VKS 313
QY 421 QSDDOIYMWANGGDPAFNIKSSFKLNNKRDRELALANGALGILKLTNNQSKRYVEV 480
DB 421 QSDDOIYMWANGGDPAFNIKSSFKLNNKRDRELALANGALGILKLTNNQSKRYVEV 480
QY 314 MKPLVANSNTVADEVERANIEFKMNNKWLFTDSGSKMTSDGINDK-----VYMLGPG 366
DB 314 MKPLVANSNTVADEVERANIEFKMNNKWLFTDSGSKMTSDGINDK-----VYMLGPG 366
QY 481 YSPLVSTLMACDEV---XXKLGDYLYLFVSYRVSRSGBRELTDNTIVGDNVAMIGY 536
DB 481 YSPLVSTLMACDEV---XXKLGDYLYLFVSYRVSRSGBRELTDNTIVGDNVAMIGY 536
QY 367 GDSLNGPHNPINETGLVLMNNLDPADLTHYSHGCIPIHPEGNNVLTSTYTNNG 421
DB 367 GDSLNGPHNPINETGLVLMNNLDPADLTHYSHGCIPIHPEGNNVLTSTYTNNG 421
QY 537 SDSLNGKYPKLLNSGVVLASVPANMKRATYSYVPAVAGHD--QVLITSYMSKDF 592
DB 537 SDSLNGKYPKLLNSGVVLASVPANMKRATYSYVPAVAGHD--QVLITSYMSKDF 592
RESULT 3
US-09-995-587A-1
Sequence 11, Application US/09995587A
Patent No. US20020127681A1
GENERAL INFORMATION:
APPLICANT: VAN HIJUM, SACHA ADRIANUS FOKKE TACO
APPLICANT: VAN GEEL-SCHUTTEN, GERITDINA HENDRIKA
APPLICANT: DIJKHUIZEN, LUBBERT
APPLICANT: RAHAOUTI, HAKIM
TITLE OF INVENTION: NOVEL FRUCTOSYLTRANSFERASES
FILE REFERENCE: BO43667-CIP
CURRENT APPLICATION NUMBER: US/09/995,587A
CURRENT FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: 09/604,958
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: EPO 00201872.9
PRIOR FILING DATE: 2000-05-25
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 789
TYPE: PRT
ORGANISM: Lactobacillus reuteri
US-09-995-587A-1
Query Match 24.7%; Score 606.5; DB 10; Length 789;
Best Local Similarity 36.2%; Pred. No. 2e-40;
Matches 156; Conservative 75; Mismatches 147; Indels 53; Gaps 17;
QY 31 QRFVQFNASAIKNIDSAKGYDKSGLNLD-LDVWDSWPLQNA-DGTAAHYGHYIVSALA 88
DB 31 QRFVQFNASAIKNIDSAKGYDKSGLNLD-LDVWDSWPLQNA-DGTAAHYGHYIVSALA 88
QY 229 RYTFEFPKASEIKNNPAAATTKQAQNTLEPLDVWDSWPLQNAQVANNMNOYQVLIAMM 288
DB 229 RYTFEFPKASEIKNNPAAATTKQAQNTLEPLDVWDSWPLQNAQVANNMNOYQVLIAMM 288
QY 89 GDPKNSDDTPLHLFYQKVDTSIDSMKNAGRYEEDMDKFPNDPYLKYQOEMSGSATLT 148
DB 89 GDPKNSDDTPLHLFYQKVDTSIDSMKNAGRYEEDMDKFPNDPYLKYQOEMSGSATLT 148
QY 289 GIP-NQNDNHITLYLLNKGNDLSHMKVGPFGYNSTAV-----SQEMSGSAVLN 338
DB 289 GIP-NQNDNHITLYLLNKGNDLSHMKVGPFGYNSTAV-----SQEMSGSAVLN 338
QY 149 KDGQVRLFTY---DYSGNPEDGGGAGANOIISTAQVNLSDPDAATLKYDGYSDHKSVEFDG 206
DB 149 KDGQVRLFTY---DYSGNPEDGGGAGANOIISTAQVNLSDPDAATLKYDGYSDHKSVEFDG 206
QY 339 SDNSIQLEFYTRVDTSDN-----NTFHOKIASATLYLTD-NGGNVSLAQVRNDYIVFE-G 390
DB 339 SDNSIQLEFYTRVDTSDN-----NTFHOKIASATLYLTD-NGGNVSLAQVRNDYIVFE-G 390

Db 1379 -----ANG-----YVKGEFVTRHGRISLYYDNGSGDQIRNRFVNAAGQMYF--DNNG 1426
QY 348 SKMTSDGINDKDYVMYMGDSDSLNCPH 374
Db 1427 YAVT-----GARTINGQH 1439

RESULT 6

US-09-815-242-5815
Sequence 5815, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 5815
LENGTH: 2076
TYPE: PRF
ORGANISM: Staphylococcus aureus
US-09-815-242-5815

Query Match 5.1%; Score 126; DB 10; Length 2076;
Best Local Similarity 23.3%; Pred. No. 0.33;
Matches 111; Conservative 57; Mismatches 204; Indels 104; Gaps 24;

QY 24 PGQONSPOKVPQFNASAIKNIDSAKGYKSGNLIDLVDWDSWPLQ--NADGFAANYHGH 82
Db 111 PSAQNQNNH---TDGNTTATEIYSNANNNDVSNNTALNV---PTKTENGSG---GHL 159
QY 83 IYSALAGPKNSDDPTPLHLFYOKVDTSIDSKNAGRVEDMDKFPVNDPYLKQYTOEWS 142
Db 160 TLKEIOEDVRHSSNNKPELY--AIAEPASNRKKRSR-----RAAPADP---NATPADP 207
QY 143 GSATLTKDGOVRLFTYDVS-----GNPEDGCTAGAGNOIIS-----TAOVNLSQPDAAATL- 191
Db 208 AAAAANGGAPVAITAPYPTPTDPNANNAGQNAPEVLSFDDNGIRQSTNRSVPVYVNVN 267
QY 192 -----KVDCVSDH-----KSVFDGSGDTVYQNIQFIDECKWISGDNHTLRDPHY 236
Db 268 NLPGETLLNGKRV-GVFSHAMVRTSMFDSGDKNKNTQAQGNVIALGRHGTDTNDHGDFFNG 326
QY 237 VED-----KGHKYLFVFEANTGTTDGYG-----DQSFNNKAYYGGSDVFF-----Q 277
Db 327 IEKALLVNNSELIFFENMTTKNGOGATNVIILKNAIDTDTIAEKTVEGGPTLRLEKVPD 386
QY 278 NEKNKLQSPKQOIASLANGALGIVELADDTYKVSVMKPLVASNTVADEVERANI--FKM 335

Db 387 NVRNLIKQFVPRKND--ITDARGIYQLKQGYKYYSFVDSIGLHSGSHVEERTDOPRAT 444
QY 336 NKKWYLFYDVS---RQSKMTSDGINDKDYV--MLGPGDGLNCPH---PINETGLVLMNN 387
Db 445 NNRKEFTVYITSLKNNNGSASLDQND--FYVQVQLPEGEVYVNNSLRKDFPSSNNGVDVN-- 501
QY 388 LDPADLTHYSHCGIPHPGNNVVLTSYMTNG-----FYPEHSHLRDLGVN 436
Db 502 -----DMNVTY-----DAANRVITIKSTGGGTANSPARLMPDKILRLRYLRVN 545

RESULT 7

US-09-815-242-12913
Sequence 12913, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 12913
LENGTH: 2186
TYPE: PRF
ORGANISM: Staphylococcus aureus
US-09-815-242-12913

Query Match 5.1%; Score 126; DB 10; Length 2186;
Best Local Similarity 23.3%; Pred. No. 0.35;
Matches 111; Conservative 57; Mismatches 204; Indels 104; Gaps 24;

QY 24 PGQONSPOKVPQFNASAIKNIDSAKGYKSGNLIDLVDWDSWPLQ--NADGFAANYHGH 82
Db 111 PSAQNQNNH---TDGNTTATEIYSNANNNDVSNNTALNV---PTKTENGSG---GHL 159
QY 83 IYSALAGPKNSDDPTPLHLFYOKVDTSIDSKNAGRVEDMDKFPVNDPYLKQYTOEWS 142
Db 160 TLKEIOEDVRHSSNNKPELY--AIAEPASNRKKRSR-----RAAPADP---NATPADP 207
QY 143 GSATLTKDGOVRLFTYDVS-----GNPEDGCTAGAGNOIIS-----TAOVNLSQPDAAATL- 191
Db 208 AAAAANGGAPVAITAPYPTPTDPNANNAGQNAPEVLSFDDNGIRQSTNRSVPVYVNVN 267
QY 192 -----KVDCVSDH-----KSVFDGSGDTVYQNIQFIDECKWISGDNHTLRDPHY 236
Db 268 NLPGETLLNGKRV-GVFSHAMVRTSMFDSGDKNKNTQAQGNVIALGRHGTDTNDHGDFFNG 326
QY 237 VED-----KGHKYLFVFEANTGTTDGYG-----DQSFNNKAYYGGSDVFF-----Q 277

Db 327 IEKALTVNSENELIEFEFTMTTKNGOGATNVIKNAEDTNDIAEKTVEGSPRLRLFKYVD 386
QY 278 NEKNLLOSPKQKQSLANGALGIYELADDTYKSVMKPLVASNTVADVEERANI--FKM 335
Db 387 NVRNLKIQFVPRKND--ITDARGIYQKDKGYYSFVDSIGLHSGHVEFRTDPAT 444
QY 336 NNKWLFTDS--RGSKMTSDGINDRDYV--MLGPGGSLNGPHN--PINETGLVLMN 387
Db 445 NKEFTVTTSKKNNSGASLDFTND--FYQVQLPEGEVYVNSLTKRDFPSSNGCYDVN-- 501
QY 388 LDPADLTHTYSHCGIPHEGNNVLTSTWTNRC-----FYPEHSHLRDLKLVN 436
Db 502 ----DMNVTY-----DAANRVITIKSTGGGTANSPARLMPKILDLRYKLRVN 545

RESULT 8
US-09-928-457-38

; Sequence 38, Application US/09928457
; Patent No. US20020164603A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA, specific proteins and peptides
; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method
; NUMBER OF SEQUENCES: 99
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (OEB)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/928,457
; FILING DATE: 2001-08-14
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 09/214,759
; FILING DATE: 199-12-10
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1981 acids amin,s
; TYPE: acide amin,
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..1981
US-09-928-457-38

Query Match 5.1%; Score 125; DB 9; Length 198;

Best Local Similarity 22.0%; Pred. No. 0.37;
Matches 95; Conservative 62; Mismatches 161; Indels 114; Gaps 23;

QY 68 LONADGTAAHYGYHIVS--ALAGDPKNSDDTPLHLFYQVGDPSIDS--WKNAGRVE 122
Db 170 IANPNCITVNGGFKFKNVGRGILITIGAPQIGKDGALTGFVROGTLTVGAAGNNDKGA-- 227
QY 123 DMDKFPVNDPYLYKYOTQEMSGSATLTKDQOVLFTIDYSGNPDGCTGAGNO--IISIA 179
Db 228 DYTGVARAVALOGLQ--GKNLAVSTGPQV--DVASGEISAGTAGTPTIALDPA 281
QY 180 QVNLQSPDPAATL--KVDVSDHKS-----VFDGSGT-----Y 211
Db 282 ALGCMYADISITLIANEKGVKNAGTLEAKOLIVTSSGRIENSGRIATTAAGTEASPT 341
QY 212 QNIQ-----QFDEGKWI-----SGDNHTLRDPHYEDGKH-----Y 244
Db 342 LSIETTEKGAAGFTISNGRIESKGLVYETGEDISLRNGAVYQNNGSRPATTVLNA 401
QY 245 LVEFANTGTGTTD-----GYGDQSFNNKAYYGGSVFPEQNEKNKLQSPKQOIAS----- 293
Db 402 LVIESKTVNNNAKGSANLSAGGRFTTINDATIQAGSSVYSTGKDELGENTRIANENVT 461

QY 294 LANGALG---IVELADDTYKSVMKPL--VASNTVADVEERANIFKNNKWLFTDSRGS 349
Db 462 LNSGISAAYIEAKQFAHIES--GKPLSLETSTVASNI-----RLNN-----GNIKG 509
QY 350 ----MTSDGINDKDYMLGPGGDSLNGPHNPINETGLVLMNLDPADLTHTYSHC--GI 402
Db 510 QALLADDDNITAKTT-----NLNTPGNLVYHGGKDLNLNVD--KDLASAASHKSDNA 560
QY 403 PHEGNNVLT 414
Db 561 AHITGTSKTLTA 572

RESULT 9
US-10-066-551-1

; Sequence 1, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the
; TITLE OF INVENTION: prevention and treatment of Neisserial infections
; FILE REFERENCE: 875, 045U1
; CURRENT APPLICATION NUMBER: US/10/066,551
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 1
; LENGTH: 2015
; TYPE: PRN
; ORGANISM: Neisseria gonorrhoeae
US-10-066-551-1

Query Match 5.1%; Score 125; DB 9; Length 2015;

Best Local Similarity 22.0%; Pred. No. 0.38;
Matches 95; Conservative 62; Mismatches 161; Indels 114; Gaps 23;

QY 68 LONADGTAAHYGYHIVS--ALAGDPKNSDDTPLHLFYQVGDPSIDS--WKNAGRVE 122
Db 170 IANPNCITVNGGFKFKNVGRGILITIGAPQIGKDGALTGFVROGTLTVGAAGNNDKGA-- 227
QY 123 DMDKFPVNDPYLYKYOTQEMSGSATLTKDQOVLFTIDYSGNPDGCTGAGNO--IISIA 179
Db 228 DYTGVARAVALOGLQ--GKNLAVSTGPQV--DVASGEISAGTAGTPTIALDPA 281
QY 180 QVNLQSPDPAATL--KVDVSDHKS-----VFDGSGT-----Y 211
Db 282 ALGCMYADISITLIANEKGVKNAGTLEAKOLIVTSSGRIENSGRIATTAAGTEASPT 341
QY 212 QNIQ-----QFDEGKWI-----SGDNHTLRDPHYEDGKH-----Y 244
Db 342 LSIETTEKGAAGFTISNGRIESKGLVYETGEDISLRNGAVYQNNGSRPATTVLNA 401
QY 245 LVEFANTGTGTTD-----GYGDQSFNNKAYYGGSVFPEQNEKNKLQSPKQOIAS----- 293
Db 402 LVIESKTVNNNAKGSANLSAGGRFTTINDATIQAGSSVYSTGKDELGENTRIANENVT 461
QY 294 LANGALG---IVELADDTYKSVMKPL--VASNTVADVEERANIFKNNKWLFTDSRGS 349
Db 462 LNSGISAAYIEAKQFAHIES--GKPLSLETSTVASNI-----RLNN-----GNIKG 509
QY 350 ----MTSDGINDKDYMLGPGGDSLNGPHNPINETGLVLMNLDPADLTHTYSHC--GI 402
Db 510 QALLADDDNITAKTT-----NLNTPGNLVYHGGKDLNLNVD--KDLASAASHKSDNA 560

OY 403 PHEGNNVLTLS 414
 DB 561 AHITGSKTLTA 572

RESULT 10

US-10-067-385-8
 ; Sequence 8, Application US/10067385
 ; Patent No. US20020110562A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Adamou, John
 ; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines
 ; FILE REFERENCE: 469201-589
 ; CURRENT APPLICATION NUMBER: US/10/067,385
 ; CURRENT FILING DATE: 2002-02-05
 ; PRIOR APPLICATION NUMBER: US/09/590,991
 ; PRIOR FILING DATE: 2000-06-09
 ; PRIOR APPLICATION NUMBER: US/60/138,453
 ; PRIOR FILING DATE: 1999-06-10
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO: 8
 ; LENGTH: 773
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 US-10-067-385-8

Query Match 4.9%; Score 120; DB 12; Length 773;
 Best Local Similarity 22.7%; Pred. No. 0.27; Indels 124; Gaps 25;

Matches 99; Conservative 59; Mismatches 154; Indels 124; Gaps 25;

OY 51 YKSGNLDLDVWDSWPLONADGTANYHGY---HVSALADDPKNSDPTPLIFYOKVG 107
 DB 91 YDDKGIILYD-----DGTLEYETEKLEDEKSKIIYGLVLSKNGHEIL-GRKS 139
 OY 108 DRSIDSWKNA---GRVEDMD-KFVPNDPYLKYOTQEMSGSAT-----LTKDGOVRLFY 157
 DB 140 NYS---KNAKYVGNVYKSIKATKYDFHSTKTFEDLYANINDIVDLAFAGDMRLFV 195
 OY 158 TYSGNPEDEGTGAGNOISTAOVNLSDP---AATLKVDYSDHKSVDGSDGDTYVYON- 213
 DB 196 KD---NDQ-----KKAELIKRMEDEKIKETKSEPRVSSVSGNVLDEBGDSLNNK 241
 OY 214 -----IQOFLDECKWISGDNHTLRDPHYVEDKHKYLVFEANTGTTDGYOGDOS 262
 DB 242 PDLTRMESGKITYSSEKQYLLKDNILIR-----KGYALKVTYVNGKIDMLEGN-- 292
 OY 263 FNNKAYYGGSDVFFQNEKKKLLQSPKKQJASLANGALGIVELADITYKS-----VVK 315
 DB 293 ---GVYSKEDI-----AKIOKANPNLRLSETTI---YADSRHVEDGRSTQSVLMS 337
 OY 316 PLVASNTVADEVERANIFKNNK-----WYLFTRDSRGSKMTSDGINKDYYMLGPGSD 368
 DB 338 ALDGRNIITVYV---FTFKMNDKGEALDKDGNLVYDS--SKLVLFGRKDKKEY---TGSD 388
 OY 369 SLNGHPNPINETGLVLMNLDPADLPTHTYSHCGIPHEGNNVLTLSYMTNRGFTYDEHHSH 428
 DB 389 REN--VEALKEDEGSMFLDTKFPVNLMSDKNYF--NPSKSNKI---YVANPEFY----- 434
 OY 429 LMDKLGAVNIKSGDTS 444
 DB 435 LRKGI-----SDKGG 444

RESULT 11

US-09-839-996-6
 ; Sequence 6, Application US/0983996
 ; Publication No. US20030009010A1
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Geme III, Joseph W.
 ; Falkow, Stanley
 ; TITLE OF INVENTION: Haemophilus Adherence and Penetration

Protein
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/839,996
 ; FILING DATE: 20-Apr-2001
 ; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/296,791
 ; FILING DATE: 25-Aug-1994

ATTORNEY/AGENT INFORMATION:
 ; NAME: Trecartin, Richard F.
 ; REGISTRATION NUMBER: 31,801
 ; REFERENCE/DOCKET NUMBER: A-59941/RET/RMS

TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:

LENGTH: 1848 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown

SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 ; US-09-839-996-6

Query Match 4.7%; Score 116.5; DB 9; Length 1848;
 Best Local Similarity 20.8%; Pred. No. 1.6;

Matches 105; Conservative 64; Mismatches 181; Indels 155; Gaps 28;

OY 25 GOONSPOKVPQFNMS--AIKNIDSAKGYDKSCGNLIDLDVWDSWPLONADGTANYHGYH 82
 DB 296 GDSGSPLEFYDREKGMWLFSGYDFWAGYNNKS-----WQENNYHFEAEKIYQOYS 348
 OY 83 IYSAL-----AGDPKNSD-----DTPHL----- 100
 DB 349 AGSLIGSNTQYWAOTGSTITGGGEPLSVLDTGKDKRPNHGSITLKGSGTLTLNNHI 408
 OY 101 -----LF---YORVGDTSIDSMKNAGRFEDMDKFPV--NDPYLKYOTQEMSGSATL 147
 DB 409 DQAGAGLPEFGDYEVKGTSDSTTWKAGVSVAD-KTYTWKVNHP--KYDRFLAKIGKGTL 465
 OY 148 TDDGOVRLFYDTSGNPF-----DG-----GTAGAQOISTAOVNL-SOPDAATLVYD 194
 DB 466 VVEGR-----GKNELGLKVGGDTVILKOKADANNKVCQAFSGVIGVSGRSTLVLLND 516
 OY 195 GVSIDHSYVDG-----GDGTYVONIQOFLDECKWISGDNHTLRDPHYVEDKHKYLV 245
 DB 517 KOYDPRNSITFFGRGRLLDNLGNSLTFDHIRN-IDGARVY--NHNMTVNSITITGES-L 572
 OY 246 VFEANTGTTDGYOGDOS-----FN--NKAYYGGSDVFFQNEKKKLLQSPK 288
 DB 573 ITNPMTITSYNEADDDHPLRIRSLPYROLVFNODNRSY-----TLKKGASTRSEPL 625
 OY 289 KOIASLANGALGIVELADITYKSWKPLVASNTVADEVERANIFKNNKMYLFTDSRGS 348
 DB 626 PONSSESENWMLYMGRTSDAARVNMNI-----NNERNNGF--NGYGEETKAT 674
 OY 349 K-----MNSDGINDDVYMLGPGDSLNGCPHNPINETGLVLMNLDPADLPTHTYSHCGI- 402
 DB 675 QNGKLVNFFNGSKDQNRFL-L-TGGTNLNGDLN-VEKGTFLFSGRPTP-----HARDIAGIS 728

OY 403 -----PH-PEGNNVYLSTMTNRCF 421
 Db 729 STKDPHFTENNEVVEDDMINRNF 753

RESULT 12
 US-10-080-505-6
 ; Sequence 6, Application US/10080505
 ; Publication No. US20030073166A1
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Geme, Joseph W.
 ; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
 ; FILE REFERENCE: A-59941-1/RFT/DOF/DHR
 ; CURRENT APPLICATION NUMBER: US/10/080, 505
 ; PRIORITY FILING DATE: 2002-02-22
 ; PRIOR APPLICATION NUMBER: US 08/296,791
 ; PRIOR FILING DATE: 1994-10-25
 ; PRIOR APPLICATION NUMBER: US 09/839,996
 ; PRIOR FILING DATE: 2001-04-20
 ; NUMBER OF SEQ ID NOS: 58
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 6
 ; LENGTH: 1848
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-10-080-505-6

Query Match 4.7%; Score 116.5; DB 9; Length 1848;
 Best Local Similarity 20.8%; Pred. No. 1.6;
 Matches 105; Conservative 64; Mismatches 181; Indels 155; Gaps 28;

OY 25 GOONSPOFKVQFNAS--AIKNIDSAGYDKSGNLIDLVDWSDPQLQADGTANYRGYH 82
 Db 296 GDSSSPLEFVVDREKGMFLGSDYFAGYNNKS-----WQENNIYKHEAEKIYQOYS 348
 OY 83 IVSAL-----AGDPKNSD-----DTPLH----- 100
 Db 349 AGSLIGSNTQYQWQATGSTTITGGEPPLSVLDTGDKDRHKGSIITLKSGTLLNNHI 408
 OY 101 -----F-----YQKVDTSIDSKNNGRFEEDMKFVP---NDPYLKYQIQEWSGSA TL 147
 Db 409 DQAGAGLEFEGDYEKGTSDSTWKGAGVSAD--GKTVTWKVNHP--KYDLAKIGKCTL 465
 OY 148 TKDQVLFYTYDSGNPE-----DG-----GTGAGNOIISTQVNL--SOPDAATLKVD 194
 Db 466 VVEEK-----GKNEGILLKYGDGVILKQKRDANNKQVAFSQVQIVSGRSTLVLLND 516
 OY 195 GVSDDHKSVEFG-----GDGTVYONIQOIFDEGKWSISGDNHTLDPHYVEDKGHXYL 245
 Db 517 KOYDPNSIYFGRGRLDLNGNSLTFDHIRN-IDDGARVY--NNHMTNTSMITITGES-L 572
 OY 246 VFEANTCTIDGYGDOS-----FN--NKAIYGGSDVFFQNEKNKLLOSPK 288
 Db 573 ITNENPTTSYNIEQDDHPLRISIPRYOLYFQNDNRSY-----TLKGASTRSEL 625
 OY 289 KOIASLANGALGIYELADDTYKSVMKPLVASNTVADEVERANIFKMNKKYLTDSRGS 348
 Db 626 PONSSEENEMWILMGRTSDAKRVNHI-----NNERNMGF---NGYFGEERKAT 674
 OY 349 K-----MTSPGINDKDYVILGPGDSLNGPHNPINETGLVILNMNLDPADLTHYSHGCI- 402
 Db 675 QNGLANTYFNKGSQDNRFLL-TGGTINLNGDLN-VEKGTLLFSGRPTP---HARDIAGIS 728
 OY 403 -----PH-PEGNNVYLSTMTNRCF 421
 Db 729 STKDPHFTENNEVVEDDMINRNF 753

RESULT 13
 US-09-839-996-4
 ; Sequence 4, Application US/09839996
 ; Publication No. US2003009010A1
 ; GENERAL INFORMATION:

APPLICANT: St. Geme III, Joseph W.
 Falkow, Stanley
 TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/839,996
 FILING DATE: 20-Apr-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/296,791
 FILING DATE: 25-AUG-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Treccartin, Richard F.
 REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELE: 910 277299
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1545 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 US-09-839-996-4

Query Match 4.7%; Score 115.5; DB 9; Length 1545;
 Best Local Similarity 20.8%; Pred. No. 1.6;
 Matches 104; Conservative 55; Mismatches 193; Indels 149; Gaps 23;

OY 25 GOONSPOFKVQFNAS--AIKNIDSAGYDKSGNLIDLVDWSDPQLQADGTANYRGYH 82
 Db 290 GDSSSPLEFVVDREKGMFLGSDYFAGYNNKS-----WQENNIYKHEAEKIYQOYS 342
 OY 83 IVSALAGDPKNSDPTPLHLYQKVDTSIDSKNNGR-----VFEDMDKFPV 128
 Db 343 AGSL-----IGSKTDYSSMSGKSTITTGGEKSLVNDLADGKDK-- 381
 OY 129 PNDPYLYQIQEWSGSA TLTKD---GOVRLFYTYDSGNPEDGCT-----GAGNOIIST 178
 Db 382 PN--HGKSVPEEGSGTLLNNINIDQAGAGLF--FEGDYEVKGTSDNTWKGAGVSVAEG 436
 OY 179 AQV-----NLSQPDAAITLVYDGVSDHKSVEFGDGDGTVYONIQOIFDEGKW----- 223
 Db 437 KTVTWKVNHPQYDLRLAKTGTLIVEGTGDKKSLKYGDGVILK--DOTNGSGHAPASV 495
 OY 224 --ISG--DNHTLDRDHYVE-----GSDVFPQNEKNKLLOSPKQOLA 292
 Db 496 GIVSGRSTLVINDKQYDPNSIYFGRGRLDLNGNSLTFPHIRNIDEGARLVHSHSKH 555
 OY 253 TTIDYQGDQSF---NNKAYV-----GSDVFPQNEKNKLLOSPKQOLA 292
 Db 556 STVITGDNLTDPDNNVSIYVKPLEDDNPYAIRQIKYGYQLYF--NEENRYYVALKKDAS 614
 OY 293 SLANGALGIYELADDTYKSVMKPLVASNTVADEVERANIFKMNKKYLTDSRGS----- 348
 Db 615 IRSEFPQNRGSSNSWLMYGTAKDAQKNA---NNHINERNRMGFGNGYFGEEGKKNNGN 670
 OY 349 -KMTSDGINDKDYVILGPGDSLNGPHNPINETGLVILNMNLDPADLTHYSHGCI----- 403

```
Db      671 LNVTFKGSSEQNRFL-TGGINLNGDLN-VOOGTLFLSGRPT- -HARDIAGISSYTK 724
QY      404 --HPEGNNVLTSTYMTNRF 421
Db      725 DSHFSENNEVVEEDWIMRNF 745

RESULT 14
US-10-080-505-4
: Sequence 4, Application US/10080505
: Publication No. US20030073166A1
: GENERAL INFORMATION:
: APPLICANT: St. Gene, Joseph W.
: TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
: FILE REFERENCE: A-59941-1/RF/DCF/DHR
: CURRENT APPLICATION NUMBER: US/10/080, 505
: PRIOR FILING DATE: 2002-02-22
: PRIOR APPLICATION NUMBER: US 08/296,791
: PRIOR FILING DATE: 1994-10-25
: PRIOR APPLICATION NUMBER: US 09/839, 996
: PRIOR FILING DATE: 2001-04-20
: NUMBER OF SEQ ID NOS: 58
: SOFTWARE: Patent version 3.1
: SEQ ID NO 4
: LENGTH: 1545
: TYPE: PRT
: ORGANISM: Haemophilus influenzae
US-10-080-505-4

Query Match      4.7%; Score 115.5; DB 9; Length 1545;
Best Local Similarity 20.8%; Pred. No. 1.6;
Matches 104; Conservative 55; Mismatches 193; Indels 149; Gaps 23;

QY      25 GOQNSPOFKVPOFNAS--AIKNIDSAKGYDKSGLIDLDVDSWPLQADGTANHYGH 82
Db      290 GDSSSPLEFYVIREKGMFLFSGSYDMAGYMKKS-----WQENNYTPEFAEKIYEDYS 342
QY      83 IVSALADGPKNSDPTPLHLEYQKVGDTSIDSMKNAGR-----VFEDMDKFV 128
Db      343 AGSL-----IGSKTDYSMSNSNGKTSTTGGKSLNVLADGKDK-- 381
QY      129 PNDYLYKQTOEMSGSATLTKD---GQVRLFTYDYSGPEDEGT-----GAGNOIIST 178
Db      382 PN--HGASVTFEGSGTLLNNIDOGAGGLE---FEGDYEVKGTSDNTTWKAGAVSVAEG 436
QY      179 AQV-----NLSOPDAATLKYDGVSDKSVFDDGDTGYONIOOFIDEKRW----- 223
Db      437 KTVWVKVHNPOYDLAKIGKGTLLIVEGTGDKGSLKVGDTVIILK--QOTNSGQHAFASV 495
QY      224 --ISG-DNHTLRDPHYVE-----DKGHKYLVEANTG 252
Db      496 GIVSGRSTLVLDNDKQVDPNSIYFGRGRLDNLGNSLTFDHIRNIDEGARLVHSTSKH 555
QY      253 TTDDYQGDQSF--NNKAYY-----GSDYFQNEKNKLLQSPKQOIA 292
Db      556 STVTTTDDNLTDPNNNSIYVKPLLEDNPPAIRQIKYQOLYF--NEENRTYVALKDKAS 614
QY      293 SLANGALGIVELADYIVKSVKPLVASNTFADEVERANIFKMNKMYLTFDSGS----- 348
Db      615 IRSFEPNREGSNNNSMLYMGTERKADAKNA-----MNIHNNRRMNGFNGFYEEEGSKNNGN 670
QY      349 -KMTSDGINDKDYMLPGGDSLNGPHNPINETGLVINMNLDPADLTHYSHCGRP----- 403
Db      671 LNVTFKGSSEQNRFL-TGGINLNGDLN-VOOGTLFLSGRPT- -HARDIAGISSYTK 724
QY      404 --HPEGNNVLTSTYMTNRF 421
Db      725 DSHFSENNEVVEEDWIMRNF 745
```

RESULT 15
US-09-912-020-381

```
: Sequence 381, Application US/09912020
: Patent No. US20020045592A1
: GENERAL INFORMATION:
: APPLICANT: Zyskind, Judith
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Trawick, John
: APPLICANT: Forsyth, R. Allen
: APPLICANT: Frieolich, Jamie M.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
: FILE REFERENCE: ELITRA.001DVI
: CURRENT APPLICATION NUMBER: US/09/912,020
: PRIOR FILING DATE: 2001-07-23
: PRIOR APPLICATION NUMBER: 09/492,709
: PRIOR FILING DATE: 2000-01-27
: PRIOR APPLICATION NUMBER: 60/117,405
: PRIOR FILING DATE: 1999-01-27
: NUMBER OF SEQ ID NOS: 485
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 381
: LENGTH: 467
: TYPE: PRT
: ORGANISM: E. Coli
US-09-912-020-381

Query Match      4.7%; Score 115; DB 10; Length 467;
Best Local Similarity 20.9%; Pred. No. 0.34;
Matches 101; Conservative 54; Mismatches 135; Indels 194; Gaps 28;

QY      34 VPOFNASAIK-----NIDSAGYDKSGLID--LDVDSWPLQAD-----GTANYHG 80
Db      5 VDQFDETFAPRDKELTSGNIADHNGVSGVPEIHSQDYVLNADLVNDRTWDTSKSYNG 64
QY      81 YHIVSALADGPKNSDPTPLHLEYQKVGDT---TSIDSMKNAGRFEEDMDKFVPNDPLYK 136
Db      65 YGIYA-----MNSDG---HLLINGNDGVNGTELDN-----SSVDVVAATGMYKV 107
QY      137 QTOEMSGSATLTKD---KDOQVRLFTYDYSN-----PEDEGTGA-----GNOI 176
Db      108 RIDNATGAGATADYDKEL--IYVNDVNSMTFSAANKRADIGATYYQAEORGNTVVLQOME 166
QY      177 --STAOVNLSDPDAAT-----LKVDGVSDKSVFDDGDTGYONIOOFIDE--GKMIS-- 225
Db      167 LTVDAVMALSPSANTMINMLEQDTV-----GTRLTNSRHGLADNGAMVSYFG 215
QY      226 ---GDNHTLRDPHYVED-----KGHLYVEANTGTTDGYQDOS----- 262
Db      216 GNFGNDGCTI---NYDQDVNGIMVGVDTKIDGNNAKIV--GAAAGFAKGDMNDRSGVDQ 271
QY      263 -----FNKKAYYGGSDVF--FQNEKKKLLQSPKQOIASLANGAL----- 299
Db      272 DSQTAIYSSAHFANNVFVDSGLSYSHFNNDLS-----ATMSNGTYVDGSTNSDA 321
QY      300 -----GIVELADYIVKSVKPLVASNTFADEVERANIF 333
Db      322 WGRFLKAGYDFKLDAGAYVTPYGSVGLFPGSGDDYQLSNMK-----VDGQSYD 370
QY      334 KMN-----NKWYLETDSRSGSKMT-----SDGINDKDYMLPGGDSL--NGPHNPIN 378
Db      371 SMREYELGVDAGYFTYSEDALEPFXKLAVYVDDSNNDNV-----NGDSIDNGTEGSAV 425
QY      379 ETGL 382
Db      426 RVGL 429
```

Search completed: June 16, 2003, 12:22:38
Job time : 71.9283 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:03:38 ; Search time 55.9426 seconds
(without alignments)
781.893 Million cell updates/sec

Title: US-09-986-682B-3

Perfect score: 2454
Sequence: 1 MNSGDYKEDYGFAHITRADM.....NIKSDTSGGENSSGQGF 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1752	71.4	489	2 F97118	levansucrase [impo
2	1568	63.9	473	2 A25040	levansucrase (EC 2
3	1552	63.2	472	2 J00802	levansucrase (EC 2
4	635	23.9	797	2 B28551	levansucrase (EC 2
5	295	12.0	428	2 H97118	levansucrase [impo
6	277.5	11.3	413	2 JC2520	beta-fructofuranos
7	258	10.5	413	2 S47527	extracellular sucr
8	255	10.4	415	2 S39195	levansucrase - Ery
9	244	9.9	423	2 S33771	levansucrase - Zym
10	243	9.9	423	2 JC2519	levansucrase (EC 2
11	237.5	9.7	430	2 JC7379	levansucrase (EC 2
12	141	5.7	1449	2 B81963	Iga-specific serin
13	139	5.7	1141	2 E89824	hypothetical prote
14	138	5.6	1165	2 T28680	fibrinogen-binding
15	133	5.4	993	2 AE1905	outer membrane sec
16	132.5	5.4	1377	2 T34632	tsh protein - Esch
17	130.5	5.3	2144	2 A97942	metalloprotease
18	129	5.3	728	2 AE0101	probable exported
19	128	5.2	1475	2 B33135	gtfB protein precu
20	127	5.2	2186	2 H89960	hypothetical prote
21	126	5.1	13055	2 T16580	hypothetical prote
22	125	5.1	759	2 AC0368	probable autotrans
23	125	5.1	2015	2 B81989	hypothetical prote
24	121	4.9	555	2 F69312	heterodisulfide re
25	120.5	4.9	3262	2 AH2137	hypothetical prote
26	120	4.9	2140	2 F95074	serine proteinase,
27	119	4.8	363	2 S43159	outer membrane por
28	118	4.8	1797	2 F69195	cell surface glyco
29	117.5	4.8	533	2 C97324	beta-xylosidase, f

30	117.5	4.8	1592	2 A38175	glucosyltransferas
31	116.5	4.7	635	2 S57714	csbp protein - Clo
32	116.5	4.7	2256	2 AD1018	large repetitive p
33	116	4.7	507	2 A40659	elastase (EC 3.4.2
34	116	4.7	582	2 S37047	beta-fructofuranos
35	116	4.7	1457	2 D81019	adhesion and penet
36	115.5	4.7	1545	2 B41859	Iga-specific metal
37	115	4.7	467	2 F64765	yalu protein - Esc
38	115	4.7	931	2 S66574	transferrin-bindin
39	115	4.7	1431	2 A45866	dextranucrase (EC
40	115	4.7	1599	2 S22737	glucosyltransferas
41	115	4.7	5291	2 F90696	hypothetical prote
42	114.5	4.7	363	2 AE0616	outer membrane pro
43	114.5	4.7	2020	2 C48399	ABC-type transport
44	114.5	4.7	3110	2 AC0116	probable virulence
45	114	4.6	730	2 F96559	hypothetical prote

ALIGNMENTS

RESULT 1

F97118 levansucrase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #extl_change 14-Sep-2001
C:Accession: F97118
R:Rolling, J., Breton, G., Omelchenko, M.V., Markarova, K.S., Zeng, Q., Gibson, R., L.
J., Daly, M.J., Bennett, G.N., Koonin, E.V., Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97118
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-489 <R>
A:Cross-references: GB:AE001437; PIDN:AAK79737.1; PID:q15024742; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1772

Query Match	71.4%; Score 1752; DB 2; Length 489;
Best Local Similarity	71.1%; Pred. No. 2.9e+11;
Matches	323; Conservative 51; Mismatches 78; Indels 2; Gaps 2;
QY	2 NSGDYKEDYGFAHITRADMKIPGQONSPOFKYPOFNASAIKIDSAGYDKSGNIDID 61
DB	30 NDMNRYKGYGVSHITRYNMSKIPMEQNDLKFYVPOFNASTLKNIASAKYDKGNLIDID 89
QY	62 VMDSWPLQVADGTAANYHGVHIVSALAGDPKNSDDPRLHLYOKVGDTSIDSKNMGVYF 121
DB	90 VMDSWPLQVADGTAANYHGVHIVSALAGDPKNSDDPRLHLYOKVGDTSIDSKNMGVYF 149
QY	122 EDMDKFVPNDPYLYKYTOEMSGSATLTKDGOVRLFTYDYSGNPEDGSGANGQIISTAOY 181
DB	150 KDSDKYVANDPYLYKYTOEMSGSATLTKDGOVRLFTYDYSGNPEDGSGANGQIISTAOY 209
QY	182 NLSQPAATLKVGVSDHKSVPFGGCGTYQNTIQGTIDEGKATISGNHTRDPHYEDXG 241
DB	210 NLSQPAATLKVGVSDHKSVPFGGCGTYQNTIQGTIDEGKATISGNHTRDPHYEDXG 269
QY	242 HKYLFVEAMTGTDTGOGQDSFNNKAYVYGGSDVFPQNEKKLL-OSPKQOIASLANGALG 300
DB	270 RKLYFEAMTGTDTGOGQDSFNNKAYVYGGSDVFPQNEKKLL-OSPKQOIASLANGALG 329
QY	301 IYELADYTVKSVKPLVASNTVADEVANIRKMNKMYLFTDSGSKWTSIDGINDKDV 360
DB	330 IYELADYTVKSVKPLVASNTVADEVANIRKMNKMYLFTDSGSKWTSIDGINDKDV 389
QY	361 YMLGPGGDSINGHNINENGLVLLNNLDPADLTHYSHGCIPIHPGCGNVVLTSTYFNG 420
DB	390 YMLGPGGDSINGHNINENGLVLLNNLDPADLTHYSHGCIPIHPGCGNVVLTSTYFNG 449

A:Title: Nucleotide sequence of levansucrase gene (levu) of *Zymomonas mobilis* ZM1 (ATCC1
 A:Reference number: S33771; MUID:93305726; PMID:818541
 A:Accession: S33771
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-423 <SON>
 A:Cross-references: GB:L08093; NID:95880348; PIDN:AAA27695.1; FID:9295699

Query Match 9.9%; Score 244; DB 2; Length 423;
 Best Local Similarity 24.6%; Pred. No. 4.7e-09;
 Matches 109; Conservative 61; Mismatches 133; Indels 140; Gaps 26;

```

QY 16 TRADMLKIPGQONSPOFKVPOFNASAIKINIDSAKGYDKSGNLIIDL-----VDSW 66
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    14 TRADAMKV--HTDDPTATMP-----TIDYDFPYMTDKYVWDWTW 50
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 PLONADGTAAHYHGHVIALAGD-----PKNSDTPPLHLFYQKVGDTSIDSKNAGV 120
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    51 PLRDINGQVSVFQGSVIFALVADRTKYGWHNRNDGARIGFYFSRSG---SNWIFGHL 106
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 FEDMDKFPVNDPFLKYQTOEMSGSATLTGKG---QVRLFYTYDSGNPEDGCTGAGNQIIS 177
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    107 LKD-----GAMP-----RSWEMSG-CTIMAPGTANSVEVFTSVNDP----- 143
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 TAQVNLSDPDAATLKVDGVSDBKSV-FDG-----GDGTVYONIQGFIDEGKWIISD 227
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    144 -----SESVPACQCKGIYIADDKSVFDFGDKVTDLFQADGLYYAD---YAEINFEW---- 190
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 228 NHTLRDPH-YVEDK-GHKYLFVEANTGTTDYGQDGSFNNKAYGSDVFPQNEKRLKQ 285
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    191 --DFRDPHFVITPKIGKTYALFEGNV-----AMERGTAVAGEEETGVP-- 231
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 SPKKQI---ASLANALGIVE-LADDTYKVSVMKPLVASNTVADEVERANIFKMNKKVYL 341
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    232 PKETETPDGARCAACAIQALNEARTKMLPLVTAFGVNDQTERRHVYFQNGLTLYL 291
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 FTDSRSGSKMTSDGINDKDYVMGLPGG-----DSLNGPHNPINETGLVL-NMNLDPADL 393
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    292 FTISHSHSY-ADGLS-----GPDGVGVFVSENGIFGYPEPLNGSLVLGNPSSQP--- 340
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 394 THTSHCGIPIHPEGNNVLTSLY 416
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    341 YQAYSHYV-----TNGLVTSFI 358
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10

levansucrase (EC 2.4.1.10) precursor - *Zymomonas mobilis*
 M:Alternate names: sucrose 6-fructosyltransferase
 C:Species: *Zymomonas mobilis*
 C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 20-Jun-2000
 C:Accession: J02519; PC2376
 R:Kyono, K.; Yanase, H.; Tonomura, K.; Kawasaki, H.; Sakai, T.
 Biosci. Biotechnol. Biochem. 59, 289-293, 1995
 A:Title: Cloning and characterization of *Zymomonas mobilis* genes encoding extracellular
 A:Reference number: J02519; MUID:95218269; PMID:7766026
 A:Accession: J02519
 A:Molecule type: DNA
 A:Residues: 1-423 <KYO>
 A:Cross-references: DDBJ:D17524; NID:9809529; PIDN:BA04475.1; PID:9809530
 A:Accession: PC2376
 A:Molecule type: Protein
 A:Residues: 1-19 <KYZ>
 C:Genetics: suc2E
 A:Gene: suc2E2
 C:Keywords: glycosyltransferase; hexosyltransferase
 F:1-19/Domains: signal sequence #status predicted <SIG>
 F:20-423/Product: levansucrase #status predicted <MAT>

Query Match 9.9%; Score 243; DB 2; Length 423;
 Best Local Similarity 24.4%; Pred. No. 5.5e-09;
 Matches 108; Conservative 60; Mismatches 135; Indels 140; Gaps 25;

```

QY 16 TRADMLKIPGQONSPOFKVPOFNASAIKINIDSAKGYDKSGNLIIDL-----VDSW 66
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    14 TRADAMKV--HTDDPTATMP-----TIDYDFPYMTDKYVWDWTW 50
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 PLONADGTAAHYHGHVIALAGD-----PKNSDTPPLHLFYQKVGDTSIDSKNAGV 120
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    51 PLRDINGQVSVFQGSVIFALVADRTKYGWHNRNDGARIGFYFSRSG---SNWIFGHL 106
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 FEDMDKFPVNDPFLKYQTOEMSGSATLTGKG---QVRLFYTYDSGNPEDGCTGAGNQIIS 177
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    107 LKD-----GAMP-----RSWEMSG-CTIMAPGTANSVEVFTSVNDP----- 143
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 TAQVNLSDPDAATLKVDGVSDBKSV-FDG-----GDGTVYONIQGFIDEGKWIISD 227
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    144 -----SESVPACQCKGIYIADDKSVFDFGDKVTDLFQADGLYYAD---YAEINFEW---- 190
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 228 NHTLRDPH-YVEDK-GHKYLFVEANTGTTDYGQDGSFNNKAYGSDVFPQNEKRLKQ 285
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    191 --DFRDPHFVITPKIGKTYALFEGNV-----AMERGTAVAGEEETGVP-- 231
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 SPKKQI---ASLANALGIVE-LADDTYKVSVMKPLVASNTVADEVERANIFKMNKKVYL 341
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    232 PKETETPDGARCAACAIQALNEARTKMLPLVTAFGVNDQTERRHVYFQNGLTLYL 291
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 FTDSRSGSKMTSDGINDKDYVMGLPGG-----DSLNGPHNPINETGLVL-NMNLDPADL 393
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    292 FTISHSHSY-ADGLS-----GPDGVGVFVSENGIFGYPEPLNGSLVLGNPSSQP--- 340
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 394 THTSHCGIPIHPEGNNVLTSLY 416
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    341 YQAYSHYV-----TNGLVTSFI 358
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 11

levansucrase (EC 2.4.1.10) - *Acetobacter xylinum*, *Acetobacter aceti* subsp. *xylinum*
 C:Species: *Acetobacter xylinum*, *Acetobacter aceti* subsp. *xylinum*
 C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 31-Dec-2000
 C:Accession: J07379
 R:Fajima, K.; Tano, T.; Kobayashi, Y.; Kohno, H.; Fujiwara, M.; Shiba, T.; Erata, T.
 DNA Res. 7, 237-242, 2000
 A:Title: Cloning and sequencing of the levansucrase gene from *Acetobacter xylinum* NCI
 A:Reference number: J07379
 A:Accession: J07379
 A:Molecule type: DNA
 A:Residues: 1-430 <TAA>
 A:Cross-references: DDBJ:AB034152
 A:Experimental source: strain NCI 1005
 C:Comment: This enzyme, having high polymerase activity, releases glucose and produce
 A:Gene: lsxA
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 9.7%; Score 237.5; DB 2; Length 430;
 Best Local Similarity 23.3%; Pred. No. 1.3e-08;
 Matches 104; Conservative 59; Mismatches 139; Indels 145; Gaps 22;

```

QY 13 AHITRADMLKIPGQONSPOFKVPOFNASAIKINIDSAKGYDKSGNLIIDL--VDSMPLO 70
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    16 SHMTIADAMKVHADDPPTTTPVYIDYNP-----VIDQVWMDTSLRA 59
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 71 ADGTAAHYHGHVIALAGDPKNSDPT-----LHLFYQKVGDTSIDSKNAGV 119
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    60 ITGETYKFNWDYVMAALVANRADTGTFVGMHNRNFAFAYIGFYISNG---ID-WTFGGR 115
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 120 VEEDMDKFPVNDPFLKYQTOEMSGSATLTG---DGOVRLFYTYDSGNPEDGCTGAGNQIIS 177
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    116 LIQKSADLRPD-----EWSGLVMRAGYKNTVDMETYSVY----- 150
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 TAQVNLSDPDAATLKVDGVSDBKSVFDG-----GGGTVYONIQGFIDEGKWIISG 226
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
    151 -TDINQSVPCSTGKIFS-NDEKVFEGFSKTVEMFSAGGVANAEEDQYFD----- 201
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```


RESULT 14
T28680
fibrinogen-binding protein homolog - Staphylococcus aureus
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE1905
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-993 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072748.1; PID:917130136; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0791

Query Match 5.6%; Score 138; DB 2; Length 1166;
Best Local Similarity 22.2%; Pred. No. 0.31;
Matches 122; Conservative 53; Mismatches 201; Indels 174; Gaps 28;

QY 3 SCDYKEYGFANHT-----TRADMLKIPGQNSPO---FKVPYNASAIKNDISA 48
DB 469 AGSQVDYGNIKKNGSTIIDONTETIKYKNSDQOLPQSNRIYDFSOYE-DVTSQPDNK 527
QY 49 KGYDKSGNLJDLVDYD-----SWPLQNDGTAA-----NYGHY 82
DB 528 KSF--SNNVATLDGPDINSAYIIKVSRYPTSGELDIAOGTSMRTTEKGYNYAGYS 585
QY 83 IVSALADPKNSDDT--PLHLFYOKVGDTSIDSKNAGRVFEDMDKEVPNDPYLYKQTOE 140
DB 586 NFIVTSNDTGGGQDGVKPEEKLY-KIGD-----YWEDVDK---DGVGTDPSKE 630
QY 141 WSGS--ATLT-KDGVRLTETTOYSGNPEDEGTAGNOIISTAOVNIQSOPA-ATLAVDG 195
DB 631 KPMANVLVTLLTPGDTTKSVRTDANGHYEFGGLKDG---ETVYKFEETPTGYLPTKVG 686
QY 196 VSDHKSVPFGDGDVYONIOQ-----FTDEGKMSGDNHTLDRPHYVEDKRGKYLVE 248
DB 687 TTIOGKSNSSSVYKINGDDMSLDGFKPKRYNLGD-----YVWEDTKKDGID 738
QY 249 ANGTGTDGYGDOSFNKAYVGGSDVFQNEKNKLLOSPKKQIASLANGALGIVELAD- 306
DB 739 ANEP-----GIKDV-----KVLKDSGTGKVGITTTDDASGKYKFTDLD 776
QY 307 --DYTVK-----SVMKPLVASNTVADEVERANI-----FKMNNK----- 338
DB 777 NGNTVFEETPAGTTPV-KNTTADDDKDSNGLTTGVYIKDADNMTLDRGYKTPKYSIGD 835
QY 339 --WTLFTDSRSGSKMTSDGINDKDYMLGPGDSLNGPHNPINETGLVLMNLD----- 389
DB 836 YVWYDSKDKDKOSTENGKIDVYVTLQNEKEVIG--TTKTDENGKRYRFENLDSGKYKI 893
QY 390 ---PADLTHYSHC--GIHPEGNNVLT-----SYMTNRGEPYEHSHLRDLGLGVNIG 439
DB 894 FEKPAGLTQVTNTTEDDKDADGGEVDYTTDHDFTLNDGXFEE----- 938
QY 440 SDTSGGENSS 449
DB 939 -DTSDDSDS 947

RESULT 15

AE1905

outer membrane secretion protein alr0791 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AE1905

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriyuchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001.

A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE1905
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-993 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072748.1; PID:917130136; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0791

Query Match 5.4%; Score 133; DB 2; Length 993;
Best Local Similarity 21.2%; Pred. No. 0.54;
Matches 99; Conservative 56; Mismatches 151; Indels 162; Gaps 21;

QY 54 SGNLIDLDVWDSWPLQNDGTAAHYHGHVYASALAGDPKNSDDPHLIFYKVD--TSI 111
DB 419 SGNLRDYSAFN-----HSTAAPSPTSLFEDYETPNFLF--LIDNNTSA 461
QY 112 DSMKNAGRVFEDMD---KEVPNDPYLYKQTOEGMSGATLTKDGOVRLFTDYSG-NPEDG 167
DB 462 QASSNLQVELQITRTVRREFVFNADY-----NGQANLTFRAMDSNGVAG 505
QY 168 G-TGAGNOIISTAOVNIQSOPDAATLKVDGVSDEKSVFEDGD--GYVQNIQPFIDECKW 223
DB 506 GTTGVAANAANGNATAFSSNTLTASTIVSPINNPIQGTGLDKLYGTANEDI----- 556
QY 224 ISGNHTLDRPHYVEDKHKYLFVEANGTGTDGYQGDQ---SFNNKAYTG--GSDVFFQ 277
DB 557 ING-----NEGNDYLFGRAGNDTLDDGEGNDYLFEGGTGNDTLDDGEGSSDLYG 604
QY 278 NEKNKLLOSPKKQIASLANGALGIVEL---ADYTVKSVMKPLVASNTVADEVER----- 329
DB 605 NEEDNII-----NGGVGNDNLDDGGTDLRGCTGNDIYTVDTGVDVIEENPE 653
QY 330 -----ANIFKMNKKWYLFDSRSGSKMTS-----DGIND----- 357
DB 654 GTDRVNSYISWTLGANLENLTILGNTIIDGTGNELDNHIIGNNAVNRLEGSDG-NDWLIG 712
QY 358 --KDYVYMLGSGGSLNRP-----NHPINETGLVLMNLDPRADLTHYSHCGIPH 404
DB 713 KDGNIDLIGGNGNRLNGETGEDTLGGLGNDVYEIDSGVYITIEAAP----- 760
QY 405 PEGNNVLTSTYMTNRGEPYEHSHLRDLGLGVNIGSDTSGGENSSGOG 452
DB 761 -AGIDTVISS-----VDWTLGVNLENLTLVGNQATLLIG 793

Search completed: June 16, 2003, 12:10:45
Job time : 58.9426 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:56:18 ; Search time 28.9037 seconds

(without alignments)
652.918 Million cell updates/sec

Title: US-09-986-682b-3
Perfect score: 2454
Sequence: 1 MNSGDYKEDYGFHITRADM.....NIKSDTSGGENSSGCGFP 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1568	63.9	473	1	SACB_BACSU
2	1552	63.2	472	1	SACB_BACAM
3	1526.5	62.2	473	1	SACB_BACST
4	636	25.9	969	1	SACB_STRSL
5	635	25.9	797	1	SACB_STRDM
6	299	12.2	415	1	SACB_RAHQA
7	281.5	11.5	584	1	SACB_ACEDI
8	274.5	11.2	413	1	INVB_ZYMO
9	274	11.0	431	1	SACB_PESH
10	269	11.0	415	1	SACB_PESG
11	255	10.4	415	1	SACB_ERAM
12	243	9.9	423	1	SACB_ZYMO
13	139	5.7	1953	1	BIGL_SALTY
14	135	5.5	1597	1	GTEL_STRDO
15	134	5.5	1476	1	GTEP_STRDM
16	117.5	4.8	1592	1	GTEP_STRDO
17	116	4.7	507	1	SEPA_STAMP
18	115.5	4.7	1545	1	IGA3_HAEIN
19	115	4.7	467	1	YAU0_ECOLI
20	115	4.7	1462	1	GTEP_STRDM
21	114.5	4.7	363	1	OMPF_SALT1
22	114.5	4.7	2003	1	YDPA_ECOLI
23	114	4.6	363	1	OMPF_SALT1
24	113.5	4.6	1849	1	IGA4_HAEIN
25	113	4.6	1325	1	YDPA_ECOLI
26	112	4.6	1165	1	YDPA_ECOLI
27	109.5	4.5	917	1	HXA3_HAEIN
28	109	4.4	687	1	INA_BACTL
29	109	4.4	810	1	GYRA_HELTP
30	109	4.4	828	1	COAT_JCDNV
31	109	4.4	1829	1	FRPC_NEITB
32	108.5	4.4	918	1	YDPA_ECOLI
33	108.5	4.4	1302	1	FRPA_NEITB

34	107.5	4.4	481	1	PRTB_ERMCH
35	107.5	4.4	1902	1	P2P_LACPA
36	107	4.4	809	1	COAT_DSDNV
37	107	4.4	811	1	COAT_GMDNV
38	107	4.4	825	1	GUN3_BACSA
39	106.5	4.3	650	1	PDR4_YEAST
40	106.5	4.3	1337	1	DEXT_STRDO
41	106.5	4.3	1902	1	P2P_LACIC
42	106.5	4.3	1902	1	P3P_LACIC
43	106	4.3	496	1	GUN3_PHAVU
44	106	4.3	1300	1	Y38_MTCPN
45	105.5	4.3	1541	1	IGA1_HAEIN

ALIGNMENTS

RESULT 1	ID	SACB_BACSU	STANDARD:	PRT:	473 AA.
AC	P05655	P70984:			
DT	01-NOV-1988	(Rel. 09, Created)			
DT	01-NOV-1988	(Rel. 09, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl				
DE	transferase) (sucrose 6-fructosyl transferase).				
GN	SACB				
OS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_TaxID=1423;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-168:				
RA	Denizot F.;				
RL	Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-168;				
RX	MEDLINE=98044033; PubMed=9384377;				
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,				
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,				
RA	Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,				
RA	Brouillet S., Brusch C.V., Caldwell B., Capiano V., Carter N.M.,				
RA	Choi S.K., Codani J.J., Connerton J.F., Cummings N.J., Daniel R.A.,				
RA	Danielot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,				
RA	Enlitan K.D., Errington J., Fabre C., Ferrari E., Foulger D.,				
RA	Filth C., Fujita M., Fujita Y., Fuma S., Gallazzi A., Galleron N.,				
RA	Glum S.Y., Glaser P., Goffeau A., Golightly E.J., Grand G.,				
RA	Giuseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,				
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,				
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,				
RA	Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,				
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,				
RA	Lee S.M., Levine A., Liu H., Masuda S., Meli C., Medigue C.,				
RA	Medina N., Mellado R.P., Mizuno M., Moser D., Nakai S., Nodack M.,				
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,				
RA	Parro V., Pohl T.M., Portetle D., Portolillo S., Prescott A.M.,				
RA	Prescan E., Pujic P., Punelle B., Rapoport G., Rey M., Reynolds S.,				
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadele Y.,				
RA	Sato T., Scanlan E., Schleich S., Schreier R., Scofield F.,				
RA	Sekiguchi J., Sekowska A., Serr S.J., Serr P., Shin B.S., Soldo B.,				
RA	Sorokin A., Taccini E., Takagi T., Takahashi H., Takemura K.,				
RA	Takenuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,				
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,				
RA	Viari A., Wambutt R., Wedler E., Wedler H., Wetzinger T.,				

```

Oy 302 VELADDYTVKSVKMPVLAASNTVADEVERANIEFKNNMNMVLTETDSRGSKMTSDGINRDVY 361
Db 316 IELMDDTYTKVKKVKKPLASNTVVDIEDERANIEFKNNKMVLTETDSRGSKMTIDGITTSDIY 375
Oy 362 MLAGGCGSLNCPNHPINDETGLVLMNMLDPADLFTYTHSCGIPHEGNGNVVLTSTYMTNKG 421
Db 376 MLEGVSNLSLNGPYRPLAKKTGLVLEKMDLDPNDVFTYTHSPAPQAKGNVVTYTSMTNKG 435
Oy 422 YPEHHSLRDLKGVNKGSTGSGENS-SCGGQ 453
Db 436 YADKQSTFPAPSEFLINIKGKKTSTVVKDSTLEBQG 468

RESULT 2
SACB_BACAM STANDARD: PRT: 472 AA.
ID SACB_BACAM
AC P21130.
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Levanucrase precursor (EC 2.4.1.10) (beta-D-fructofuranosyl
  transferase) (Sucrose 6-fructosyl transferase).
GN SACB.
OS Bacillus amyloliquefaciens.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1390;
  [1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 23844;
RX MEDLINE=91092506; PubMed=2265762;
RA Tang L.B., Lenstra R., Borchert T.V., Vasantha N.;
RT "Isolation and characterization of levanucrase-encoding gene from
  RL Bacillus amyloliquefaciens.";
  Gene 96:89-93(1990).
CC -i- CATALYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =
  CC glucose + ((2,6)-beta-D-fructosyl)((N+1)).
  CC -i- SUBCELLULAR LOCATION: Secreted.
  CC -i- INDUCTION: BY SUCROSE.
  CC -i- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
  CC -----
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@isb-sib.ch).
  CC -----
  CC EMBL: X52988; CAA37179.1; -
  DR PIR: J00802; J00802.
  DR PIR: S11739; S11739.
  DR InterPro: IPR003469; Glyco_hydro_68.
  DR Pfam: PF02435; Glyco_hydro_68; 1.
  DR Transferrase; Glycosyltransferase; Signal.
  FT SIGNAL 1 29
  FT CHAIN 30 472 LEVANSUCRASE.
  SO SEQUENCE 472 AA; 52859 MW; F38592D272677E7D CRC64;

Query Match 63.2%; Score 1552; DB 1; Length 472;
Best Local Similarity 65.3%; Pred. No. 1,de-92;
Matches 296; Conservative 54; Mismatches 85; Indels 18; Gaps 5;

Oy 2 NSGGRKEDYGEVNIHTRDMLKIPCOQNSPQFKYVQFNASAIKNIDSKAGYDKSNLDD 61
Db 33 NQKAYKTEGVSHTRHDMLOIPKQOQNEKYVQFOPOSTKTNIESAG-----LD 83
Oy 62 VWDSMPLOMAGCTAANYHGYHVALAGDPKNSDDTFLHLFYOKVCGTSDSMKNAGRVF 121
Db 84 VWDSMPLOMAGCTAANYHGYHVALAGSPKADADDTSTIYFYQVGVGNSIDSMKNAGRVF 143
Oy 122 EDMKRFVNDPDPYLYKQTQEWSGSATLTTRKDQVRLFTYTDYSGNPEDGGTGAGNQILSTRAV 181

```

Db	144	KDSDFDANDPLKQDQJEMSGSASFISDGRIRLFTDYDSGK-----HYGQSULTAQY	197
Qy	182	NLSQPDATLKVDCYSDHKSVFDDGDCGYIYQNIQOIFIDEGKATISGDNHTLRDPHYEDKG	241
Db	198	VS:KSD-FTLKNINGVEDHKHTIFD-GDGKYYQNVQOOFIDEGNATISGNNHLRDPHYEDKG	255
Qy	242	HKYLFEEANTGTTDYOQDOSFENNRKAYYGSDVPEONEKNKTLQSPBPKGIASLANGALGI	301
Db	256	HKYLFEEANTGTEGNGEYQGEESLFENKAYYGCGGNFRKEBSQKTLQSSAKKRDALENALGI	315
Qy	302	VELADDTYVYKSVMKRLVASNTYAYADEVERANITKANNKWTLLFTDSRGSKATSDGINDKDY	361
Db	316	IELANDNYLLKVKMKELITSNTYIDIEIRANVYKMKMGKWTLLFTDSRGSKMTIDGINSNDY	375
Qy	362	MLPGSGDLSHPHNEINETGLVANNLRPADLFTHYHSHCGIPIHPDEGNNTVLTSYMTNRGF	421
Db	376	MLGYVSNLSLTGPKRLNKTGLVLQMGDLDPNDVTFTYSHFAVYQAKGNVNYITSYMTNRGF	435
Qy	422	YPEHSHLRDLGVNIKSDTSISGGENS-SGQGO	453
Db	436	FEDKATGPGSPFLMIKGNKTSVYVNSILSQGO	468

```

1  RESULT 3
2  SACC_ BACST
3
4  ID SACC_BACST STANDARD; PRT; 473 AA.
5  P94468.
6  DT 15-DEC-1998 (Rel. 37, Created)
7  DT 15-DEC-1998 (Rel. 37, Last sequence update)
8  DT 15-DEC-1998 (Rel. 37, Last annotation update)
9  DE Levanucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
10 transferase) (Sucrose 6-fructosyl transferase).
11 SABC OR SURB.
12 OS Bacillus stearothermophilus.
13 OC Bacteria; Firmicutes; Bacillales; Geobacillus.
14 OX NCBI_TaxID=1442;
15 [1]
16 RP SEQUENCE FROM N.A.
17 RC STRAIN=ATCC 12960;
18 MEDLINE=98007870; PubMed=9349714;
19 Li Y., Triccas J.A., Ferenc T.;
20 RA "A novel levanucrase-levanase gene cluster in Bacillus
21 stearothermophilus ATCC12980 ";
22 Biochim. Biophys. Acta 1353:203-208(1997).
23 CC -I- CATALYTIC ACTIVITY: Sucrose + [(2,6)-beta-D-fructosyl].(N) =
24 glucose + [(2,6)-beta-D-fructosyl].(N+1).
25 CC -I- SUBCELLULAR LOCATION: Secreted.
26 CC -I- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
27 -----
28 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
29 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
30 CC the European Bioinformatics Institute. There are no restrictions on its
31 CC use by non-profit institutions as long as its content is in no way
32 CC modified and this statement is not removed. Usage by and for commercial
33 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
34 CC or send an email to license@sib-sib.ch).
35 -----
36 CC EMBL; U34874; AAB9711.1; -
37 DR InterPro; IPR003469; Glyco_hydro_68.
38 DR Pfam; PF02435; Glyco_hydro_68.1.
39 FT Transferrase; Glycosyltransferase; Signal.
40 FT SIGNAL 1 29 BY SIMILARITY.
41 CHAIN 30 473 LEVANUCRASE.
42 SEQUENCE 473 AA; 52854 MW; 12F7A272EC2B771E CRC64;

```

	Query Match	62.2%	Score 1526.5	DB 1	Length 473
	Best Local Similarity	65.2%	Pred. No. 5.9e-91		
	Matches 288	Conservative 52	Mismatches 85	Indels 17	Gaps 4
Qy	2	NSGDIEDYGAHTRADMLKIPGQONSPOEFKVPQFNASAIKINDASAGYDKSGMLIDLD	61		
Db	33	NQRYKEVYGSHTIRHMDLIPQOQNKKEKQVDFSDSTKTNISSAG-----LD	83		

```

QY 62 VMSWNPLOWNADGAAANYHGHYHVSALADPKNSDDTPLHLFTYQKVGSTSIDSKMNGRVF 121
Db 84 VMSWNPLOWNADGVAANYHGHYHVSALADPKNADDTSLYMYQKVGSTSIDSKMTPDVRVF 143
QY 122 EMDDKFVPPNDPYLKYTOEWSGSAFLTKDGQVRLEPTYPTSGNPPEDGSGTGAONLISPAQY 181
Db 144 KSDDKFKDADDSLAKODTOEWSGSAFLTDGKRLRFLPYTDSK-----HYGKOTLTTAQY 197
QY 182 NLSQPDATLTKYDGVSDHKSVFDGGDGTGYVYONIQOFLDEGKMISGDNHLLRDPHYEDKG 241
Db 198 NWSASP-SSLININGVEDYKSIFF-GDSKTYQWVOQFIDEGWYSSGDHMLTRDPHYEDKG 255
QY 242 HKTLVEANVTGTTDGVQDSQSFNNKAYYGVGSVPFQNKKNLLOSPKKQJLASLANGLGI 301
Db 256 HKTLVEANVTGTTDGVQDSQSFNNKAYYGVGSVPFQNKKNLLOSPKKQJLASLANGLGI 315
QY 302 VEIADDTYKVSVMKPLVASNTVADAEVERANIEFKMNKKYVLTDSRGSKMTSDGINDADYV 361
Db 316 IELNDOTLTKKWKPLIASNTVYTDLEIRANNEYKMNKKRYLSTDSRGSOMNTIDGITSMDIY 375
QY 362 MLCRGDDSLNGPHNPINETGLVLANNLDPADLTHTYSHGCIPIHDEGNVYLTSYMTNRCF 421
Db 376 MLCYVNSNLTGPPKPLNKTGLVLKDDLPDNDVTFYSHFAVQATGNVNYTISYMTNRCF 435
QY 422 YPEHSHLKDLCVNTKSGSDTS 443
Db 436 YADKQSTFAPSPFLNIOGKRTS 457

```

```

RESULT 4
SACB_STRSL          STANDARD:          PRT:          969 AA.
ID  SACB_STRSL
AC  Q55242.1
DT  15-DEC-1998 (Rel. 37, Created)
DT  15-DEC-1998 (Rel. 37, Last sequence update)
DT  15-DEC-1998 (Rel. 37, Last annotation update)
DE  Levanucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
DE  transferase) (Sucrose 6-fructosyl transferase).
GN  FTF.
OS  Streptococcus salivarius.
OC  Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC  Streptococcus.
OX  NCBI_TaxID=1304;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 25975;
RX  MEDLINE=93322332; PubMed=8331080;
RA  Ratchesam C., Giffard P.M., Jacques N.A.;
RT  "The cell-bound fructosyltransferase of Streptococcus salivarius: the
RT  carboxyl terminus specifies attachment in a Streptococcus gordonii
RT  model system.";
RL  J. Bacteriol. 175:4520-4527(1993).
CC  -I- CATALYTIC ACTIVITY: Sucrose + [(2,6)-beta-D-fructosyl]}(N) =
CC  glucose + [(2,6)-beta-D-fructosyl]}(N+1).
CC  -I- SUBCELLULAR LOCATION: SECRETED. CELL-WALL BOUND.
CC  -I- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL collaboration -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement. (See http://www.isb-sib.ch/announc/
CC  or send an email to license@isb-sib.ch).
```

DR	EMBL, L08445; AAA71925.1; -	
DR	InterPro; IPR003469; Glyco_hydro_68.	
DR	Plam; PF02435; Glyco_hydro_66; 1.	
DR	TIGRFAMs; TIGR01168; YSKR_signal; 1.	
KW	Transferase; Glycosyltransferase; signal; Cell wall.	
FT	SIGNAL	1 ? POTENTIAL.
FT	CHAIN	? 969 LEVANSUCRASE.
Q3	SEQUENCE	969 AA; 103983 MW; D39B5B332AC735A CRC644

Query Match 25.9%; Score 636; DB 1; Length 969;
 Best Local Similarity 36.7%; Pred. No. 2e-33;
 Matches 168; Conservative 78; Mismatches 128; Indels 84; Gaps 23;

26 QONSQFQKPPQNASIKKIDSAKGYD-KSGNLIDLDVDSMPLQNA-DGTAAHYHGYH 83
 249 KQDS-KLAVYFKADITIMMPAFNTVDAQTKEEDLSDMPVDAAKSGVSNMNGYOL 307
 84 VALAGDKPKNSDDTPLHLFYOKVGPSTIDSMKNAGRVF-----EDMDKRVPPNDPYLYQT 138
 308 VISMAGAP-NKSNHIIYLLRYKYGNDFTHKMAGPIEGYNALED-----D 352
 139 QEMSSGATITKDGQVRLFTYDYSGNPEDGTGAGNOIISTAOVNI SQPDAATLKVDGVS 198
 353 QOMSGATVNSDGSIOLYTK---NDTSGCKLNMWOLAS-ATLNLAVENDEVVAKSVEND 408
 199 HKSVPDGGGTYQYQIQDID-----EGKMSIDNITLDRPHYEDKGRHLYVEAN 250
 409 H--LLEGGNYHYQSTPKFMSFTDDHNDHNPDRDNTCLDRPHILENGSRYLFEESN 466
 251 TGTDTGOGDQSEFNKAYYGSDFV---FON-EKNKLQSPKQIATSLANGALGIVELA 305
 467 TG-DEYQGEKQITKMSNYGGDAFNLKSFNLIVNKKLYN---LASMANGSIGILKLD 521
 306 DTY---TVKSYMKPLVASFTVADEVERANIFKMNKNWYLFDSRSGKMTS-----DG 354
 522 DNEKNPSVALTYPTLPTSHMTVDEVERPSVVKMGKGYLFETASRIKSTDAEGTVARBA 581
 355 INDKPYVMKPGGDSLNGHNPINETGLVLMNLDPAD-LTHYSHCGIPHEGNN--VY 411
 582 VSD-DVVMGLPVSDSLRGKIRPLNSGVVLTASV-PADWRTSTSYSAVP-VEGSSDTLL 638
 412 LFSYMTNRGEYFDEHSHLRDLKLVNKGSDTSGGENSS 449
 639 VTSYMTNRG-----GI-----AGAENKS 656

RESULT 5
 SADB_STRMU STANDARD; PRT; 797 AA.
 ID SADB_STRMU STANDARD; PRT; 797 AA.
 AC P11701;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
 transferase) (Sucrose 6-fructosyl transferase).
 GN FTF.
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX STRAIN=GS-5;
 RX MEDLINE=88115184; PubMed=2828325;
 RA Shitoto T., Kuramitsu H.K.;
 RT "Sequence analysis of the Streptococcus mutans fructosyltransferase
 gene and flanking regions."
 RL J. Bacteriol. 170:810-816(1988).
 CC -i- CATALYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =
 glucose + ((2,6)-beta-D-fructosyl)(N+1).
 CC -i- SUBCELLULAR LOCATION: Secreted.
 CC -i- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL: M18954; AAA88584.1; -
 DR PIR: B28551; B28551.
 DR InterPro: IPR003469; Glyco_hydro.68.
 DR Pfam: PF02435; Glyco_hydro.68; 1.
 KW Transferase; Glycosyltransferase; Signal.
 FT SIGNAL 1 34
 FT CHAIN 35 797 LEVANSUCRASE.
 SQ SEQUENCE 797 AA; 87574 MW; 6DDE057465DBA042 CRC64;

Query Match 25.9%; Score 635; DB 1; Length 797;
 Best Local Similarity 39.0%; Pred. No. 1.8e-33;
 Matches 159; Conservative 74; Mismatches 135; Indels 40; Gaps 19;

31 QKRVQFNASATKKNIDSAKGYD-KSGNLIDLDVDSMPLQNA-DGTAAHYHGYHVALA 88
 214 RYALDYFNKAKIKKMAATRDAGTQQLADLDVMSWPVDAAKTEVINMNGYOLVAMM 273
 89 GPKNSDDTPLHLFYOKVQVDTSIDSMKNAGRVFEDMDKRVPPNDPYLYQTQOMSGSATLT 148
 274 GIP-MTNDHIIYLLNKYGDNFDMKNAGSLFG-----YNETPL---TQEMSGSATVN 323
 149 KDGQVRLFTYDYSGNPEDGTGAGNOIISTAOVNI--SQPDAATLKVDGVSDFKSVFPG 206
 324 EDGSLQLEFTTKVDTSDKN---SNNQRLATATVNLGFDQDVRILSYE--NKKVLTPEGV 377
 207 DGTVYQNIQOFTDEGKMSISGDNHNLDRPHYEDK-GHKLYFEANTGTTDGYQDQSEFN 265
 378 MAYHYQSYQW--RSTFTGADNIARDPHVIDEDNGDRLYVEASTG--TENYQGEDQIYN 434
 266 KAYYGSDFVFPNEKKNKLQSPK-QIASLANGALGIVELADYIVKSV---MKPLVASN 321
 435 FTVYQSSASVNVKSLFRFLDDQDMYNRASWANAAGILKLDKDKTPVEDQFYTPLLST 494
 322 TVADEVERANIFKMNKNWYLFDSR---GS-----KMSDGIKNDVYMLGPGDSLNGP 373
 495 MVSDELEERNVYKLDQKYLFTASRLNHGNSNDANMKANEEVGD--NVMLGIVSSQQLTNG 553
 374 HNPINETGLVLMNLDPAD-LTHYSHCGIP-HPEGNNVLTSTYMTNR 419
 554 YKPLNNSGVVLTASV-PADWRTATYSYVAVPAGSSDTLLMAYMTNR 600

RESULT 6
 SADB_RAHAO STANDARD; PRT; 415 AA.
 ID SADB_RAHAO STANDARD; PRT; 415 AA.
 AC 054435;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Levansucrase (EC 2.4.1.10) (Beta-D-fructofuranosyl transferase)
 DE (Sucrose 6-fructosyl transferase).
 GN SADB.
 OS Rahella aquatilis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Rahella.
 OX NCBI_TaxID=34038;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC 33071;
 RX MEDLINE=99127094; PubMed=9928133;
 RA Song K.B., Seo J.W., Kim M.G., Rhee S.K.;
 RT "Levansucrase of Rahella aquatilis ATCC33071. Gene cloning,
 expression, and levans formation."
 RL Ann. N.Y. Acad. Sci. 864:506-511(1998).
 CC -i- CATALYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =
 glucose + ((2,6)-beta-D-fructosyl)(N+1).
 CC -i- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U91484: AAC36458.1: -;
DR InterPro: IPR003469: Glyco_hydro_68.
DR Pfam: PF02435: Glyco_hydro_68; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 415 AA: 45939 MW: E02828813D13A74 CRC64;

Query Match 12.2%; Score 299; DB 1; Length 415;
Best Local Similarity 26.2%; Pred. No. 2.8e-12;
Matches 116; Conservative 60; Mismatches 141; Indels 126; Gaps 24;

OY 16 TRADMLKIPGQONSPOFKVPOFNASAIKINDSAKDYKSGNLIDP-----VMSW 66
DB 12 TRADALKV--NENPPTTQP-----IVADPFVNSDEVITWDTM 48

OY 67 PLQNAADGTAAHYHGYHVSALAGDPKNSDPTPLHLYOKVSDTIDS-WKN----- 116
DB 49 PLRSIDGTIVSVDSGVSIFTLTAQRNNNSE---YLDAGNYDITSDMNNRHRARICY 104

OY 117 -----AGRFEDMDKFPVNDPLKTYOTQEMSSS-ATLTKDGOVRLFTYDSGNP 164
DB 105 WYSRTGKDWIFGGRVMAE-----GVSPTSREMACTPILLNEDGDIDLYTCVT--- 152

OY 165 EDGGTAGCNOIITSTAOVNLSPDAATLKYDGVSDHKSVFDDGDTVTYONIDQFIDEKWI 224
DB 153 -----PGATIAKVRGKVLTSSEGYTLA--GFENEKSLF-SADGYVYOTESO---NPTYN 200

OY 225 SGNHTLTPHYVEDKGRKYLVEFANTGTCTDGYODGOSFNKAYVYGSDFEQNEKNKL 284
DB 201 FRDPSPTFDPP-----DGKLYMVEGNVA-----GSHVIGKQMGTL- 240

OY 285 QSPKQOIASLNGALGIYELA-----DOYTVKSVMKPLVASNTVADEVERANIFKANN 337
DB 241 -PGRHVGNAKRYOAGCIGMAVAKDLSGDEW---EILPLVTAVGVNDQTERPHRFVODG 296

OY 338 KMYLTDSRSGKMT-SDGINDKD-VYMLGPGDSLNGPHNPINETGLVNNHLPADLTH 395
DB 297 KYVLETFISH-KFTYADGLTGRPDGY--GLSLDNLGTGYSPRNGSLVGLGN--PPSQPQ 350

OY 396 TYSHCGIHPPEGNNVLTSTMTN 418
DB 351 TYSHCVMP-----NGLVTSFIDN 368

RESULT 7
SACB_ACEDI STANDARD: PRT; 584 AA.
AC 043998;
DT 15-DEC-1998 (Rel. 37; Created)
DT 15-DEC-1998 (Rel. 37; Last sequence update)
DT 15-JUN-2002 (Rel. 41; Last annotation update)
DE Levansucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl
DE transferase) (Sucrose 6-fructosyl transferase).
GN LSDA.
OS Acetobacter diazotrophicus (Gluconacetobacter diazotrophicus).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
CC Gluconacetobacter.
CC NCBI_TaxID=33996;
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 52-61.
RC STRAIN-SRT4 / CBS 550.94;
RX MEDLINE=96253999; Pubmed=8704949;
RA Arieta J., Hernandez L., Coego A., Suarez V., Balmori E.,
RA Mendez C., Pelt-Glatron M.-F., Chambert R., Selman-Housein G.;
RT "Molecular characterization of the levansucrase gene from the
RT endophytic sugarcane bacterium Acetobacter diazotrophicus SRT4.";
RL Microbiology 142:1077-1085(1996).
CC -i- FUNCTION: RELEASES FRUCTOOLIGOSACCHARIDES AND LEVAN, A HIGH-
CC MOLECULAR-MASS FRUCTOSYL POLYMER, FROM SUCROSE. IT ACTS MORE AS A
CC SUCROSE HYDROLASE THAN AS A FRUCTAN POLYMERASE.
CC -i- CATALYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =

CC glucose + ((2,6)-beta-D-fructosyl)(N+1).
CC -i- SUBCELLULAR LOCATION: Secreted.
CC -i- PTM: THE N-TERMINAL IS BLOCKED.
CC -i- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L41732: AAB36606.1: -;
DR InterPro: IPR003469: Glyco_hydro_68.
DR Pfam: PF02435: Glyco_hydro_68; 1.
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 30
FT PROPEP 31 51
FT CHAIN 52 584 LEVANSUCRASE.
SQ SEQUENCE 584 AA: 63525 MW: CBFBA139AD0B8CE CRC64;

Query Match 11.5%; Score 281.5; DB 1; Length 584;
Best Local Similarity 26.4%; Pred. No. 5.8e-11;
Matches 120; Conservative 59; Mismatches 184; Indels 91; Gaps 26;

OY 5 DYKEEYGFANHTRAAMKI-----PGQONS-GRKVPQFNAS-AIKNDSAKGYDK 53
DB 79 DPQSDP-TARWTRADALDIKAHSDATVAAGONSIPQAOLTPMPADPFVIN----- 128

OY 54 SGNLIDLVMSWPLQNAADGTAAHYHVSALAGDPK-----NSDPTPLH-----LEYOKV 106
DB 129 -----PDVWVMDPTWTLIDKHADQFSTNGWEVIFCLTADNAGYGPDRVHARIGFTYRA 184

OY 107 GDTST-----DSKNAGVFEED--MDKFPVNDPLKTYOTQEMSGSATLTR--DGQVRLFY 157
DB 185 GIPASRRPVNGGWTPYGHILFPDGASAOVYAGQTY--NOAEMSGSSRLMOIHGNTVSFVY 242

OY 158 TDYSGNPNPDGGGAAN-----QIISTAOVNLSPDAATLKYDGVSDHKSVFDDGDTVTYON 213
DB 243 TDVAFNRD--ANANNITPPOAILITQIGRIHADENHWFGEFNAHTPLLO-PDGLVLQN 298

OY 214 IQQFIDECKWISGDHNTLRDPHYVEDKCH---KYLVEFANT-----GTTDGYODGOSFNNK 266
DB 299 GAQ-----NEFFNRDPPTFEPRKHPGVNYVFEAGNTAGQGVANCTEADLGFR-- 347

OY 267 AYYGSDVFPQNEKN-KLLQSPKQOIASLANGALGIVELADDYTK--SVMKPLVASMTV 323
DB 348 -----PNDPNAETLQEVLDGSAVYQKANKILA-IATDSTLSKKKPLSPLISANCY 396

OY 324 ADEVERANIFKNNKMYLFTDSRSGKMTSDGINDKD-VYMLGPGDSLNGPHNPIN-ETG 381
DB 397 NDQTERPOVYLLNGKYVYIFITISHRTFAA-GVDGPDGY--GFVDDGIRSDPQPNVYSG 453

OY 382 LVNLNMLDPADLTHTYSHCGIHPPEGNNVLTSTY 415
DB 454 LTMG---NPTDLNTAAGDPDPSPDQNRARQSY 484

RESULT 8
INVB_ZYMO STANDARD: PRT; 413 AA.
AC 060115; 060117; 060125;
DT 15-DEC-1998 (Rel. 37; Created)
DT 15-DEC-1998 (Rel. 37; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Extracellular sucrose (EC 3.2.1.26) (Beta-fructofuranosidase)
DE (Invertase) (Saccharase) (Protein B46).
GN SACC OR INVB OR SUCR3.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
CC Zymomonas.
CC NCBI_TaxID=542;

QY 172 GNOIITSAQVNLSDPATLTKVDG---VSDH-----KSYFDGDDGTYYONIOFI 218
 Db 169 -----PGAAIAKVRGRIVTSODGVELKDFQVKRLF-ADGYYIYTEAO-- 211
 QY 219 DEGWISGDNHTLDPHYVEDKGHKYLVFEANTGTGOGDOSFNKKAYVGGSDVEFQ 278
 Db 212 -NSSMNRDPSFIDP-----NDGKLYWFEENVA---GERGSH-----VGAELGPVP 257
 QY 279 EKNKLLSPKQIATSLANGALGIVELADDTYK--SVMKPLVASTVADYERANIFKM 336
 Db 258 PGHEHETGARGVY-----GCGILA-VAKDLSGEEMELLPLVTVAVGNDOTERPHVFO 311
 QY 337 NKWYLFDSRGSKMT-SDGINDKD-VYMLGPGDSLNGPHNPINETGLVLMNLDPADLT 394
 Db 312 GKYYLFTISH--KFTYADGLTGPDCYV--GFVGEHLFGPRPMNASGLVLGN--PPDQPF 365
 QY 395 HTYSHCGIPHEGNNVLTSTM 416
 Db 366 QTYSHCVMP-----NGLVTSFI 382

RESULT 10

SACB_PSESG STANDARD: PRT: 415 AA.
 ID SACB_PSESG
 AC 052408;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Levanusucrase (EC 2.4.1.10) (beta-D-fructofuranosyl transferase)
 DE (Sucrose 6-fructosyl transferase).
 GN LSC.
 OS Pseudomonas syringae (pv. glycinea).
 OC Bacteria: Proteobacteria; gamma subdivision: Pseudomonadaceae;
 CC Pseudomonas
 OX NCBI_TaxID=318;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=PG4180;
 RA MEDLINE=98394981; PubMed=9726857;
 RT Hettwer U., Jaekel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.:
 "Cloning, nucleotide sequence, and expression in *Escherichia coli* of
 levanusucrase genes from the plant pathogens *Pseudomonas syringae* pv.
 RT *glycinea* and *P. syringae* pv. *phaseolicola*."
 RL Appl. Environ. Microbiol. 64:3180-3187(1998).
 CC -1- CATALYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =
 CC glucose + ((2,6)-beta-D-fructosyl)(N+1).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF037443; AAC36056.1;
 DR InterPro: IPR003469; Glyco_hydro_68.
 DR Pfam: PF02435; Glyco_hydro_68; 1.
 KW Transferase; Glycosyltransferase.
 SQ SEQUENCE 415 AA; 45844 MW; 839B686AC80610CF CRC64;

Query Match 11.0%; Score 269; DB 1; Length 415;
 Best Local Similarity 25.1%; Pred. No. 2,4e-10;
 Matches 111; Conservative 66; Mismatches 136; Indels 130; Gaps 24;

QY 16 TRADMLKIPGQNSPOFKVPQFNASAIKINDSAGYKSGNLDLDVWDSWFLONADGTA 75
 Db 12 SRADALKV--NENDPTTQPLVSPDPFVMSDF-----VFIMDTMFLELDGTV 57
 QY 76 ANHGHIIVSALAGDPKNSDDTPIHL-----FYQVGDTSID 112

Db 58 VSVNGMSVITLTAD-RHPDD-PQYVGANGRYDIKRDMEHRHGRARMCWYSRTG-----K 111
 QY 113 SMKAGAVFEDMDKFVNDPLKQTOBMSGSAITLTD-GVRLFTYDYSGNPDGCTGA 171
 Db 112 DMIFGGRVME-----GVSPTRBMAGTPLVLLNDKDDIDILYXCVCV----- 152
 QY 172 GNOIITSAQVNLSDPATLTKVDG-----VSDHKSVDGDDGTYYONIOFI 218
 Db 153 -----PGAAIAKVRGRIVTSODGVELKDFTEVKRLF-ADGKYIYTEAO-- 195
 QY 219 DEGWISGDNHTLDPHYVEDKGHKYLVFEANTGTGOGDOSFNKKAYVGGSDVEFQ 278
 Db 196 -NSTNMNRDPSFIDP-----NDGKLYWFEENVA---GERGSH-----VGAELGPVP 241
 QY 279 EKNKLLSPKQIATSLANGALGIV--ELADDTYKSVMKPLVASTVADYERANIFKM 335
 Db 242 PGHEHETGARGVY-----GCGILAVARDLSGDEM-BILPLVAVGNDOTERPHVFO 294
 QY 336 NKWYLFDSRGSKMT-SDGINDKD-VYMLGPGDSLNGPHNPINETGLVLMNLDPADL 393
 Db 295 GKYYLFTISH--KFTYADGVTGPDCYV--GFVGEHLFGPRPMNASGLVLGN--PPAOP 348
 QY 394 HTYSHCGIPHEGNNVLTSTM 416
 Db 349 QTYSHCVMP-----NGLVTSFI 366

RESULT 11

SACB_ERWAM STANDARD: PRT: 415 AA.
 ID SACB_ERWAM
 AC 046654;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Levanusucrase (EC 2.4.1.10) (beta-D-fructofuranosyl transferase)
 DE (Sucrose 6-fructosyl transferase).
 GN LSC.
 OS *Erwinia amylovora*.
 OC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae;
 CC *Erwinia*.
 OX NCBI_TaxID=552;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EA7/74;
 RA Geier G., Geider K.K.:
 RT "Characterization and influence on virulence of the levanusucrase gene
 RT from the fireblight pathogen *Erwinia amylovora*."
 RL Physiol. Mol. Plant Pathol. 42:387-404(1993).
 CC -1- CATALYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =
 CC glucose + ((2,6)-beta-D-fructosyl)(N+1).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X75079; CAAS2972.1;
 DR InterPro: IPR003469; Glyco_hydro_68.
 DR Pfam: PF02435; Glyco_hydro_68; 1.
 KW Transferase; Glycosyltransferase.
 SQ SEQUENCE 415 AA; 46408 MW; 4FF564F6E0607FEB CRC64;

Query Match 10.4%; Score 255; DB 1; Length 415;
 Best Local Similarity 25.4%; Pred. No. 1.9e-09;
 Matches 115; Conservative 64; Mismatches 141; Indels 132; Gaps 29;

QY 9 DYGFANI--TRADMLKIPGQNSPOFKVPQFNASAIKINDSAGYKSGNLDLD----- 61

```

Db      3  DYNRPPLTMTRADALKV - HEDDPPTTQO-----VIDIARPVMS 39
Qy      62  ----VWDSMPLONAGSTANHYGHVLSVLADGPKSSDTPHL-----1000
Db      40  EEVFLTMDPLMEDEFGEIISVNGWCIIFFLTAD-RNTDNPQOFEDENGNYDITRDWEDRHG 98
Qy      101  ----LFYOKVSDTSDSKWNKRNARFVEDBDRKVPNDPFLKYQFOEWSGSATLEKD-GOVR 154
Db      99  RARICWVSRG-----KDMIFGGRVW -AEGVAPF-----TRENAQPIILINDGDID 145
Qy      155  LFYTYSGNPEDEGTTGA--GNQIISTAOVNLSDPDAATLKVDGVSDBKSVFDGSGDTYV 211
Db      146  LYTTCVT-----PGATIAKVRGKIYVSDQSVSL-----EGFOQVTSLF-SADGTYI 190
Qy      212  QNIQOFIDEGKNIISGDNHTLRDPHYEKK--GHKIYVEFANIGTTPGQGDGSFNNKAYV 269
Db      191  QTEEO--NAFW-----NFRDPSPEIDNDCKLYLVEGNAV--GPRGSHEI-TQAEW 237
Qy      270  GGSVDYFONEKKLLQSPKKQJASLANGALGIVELADDDTVK--SVYMKPLVASNTVADEV 327
Db      238  GNVPFGYEDYGGAKKQQA-----GCVGLA-VAKDLSGSWMOLLPLITAVGVNDQY 286
Qy      328  ERANIFKNKKWYLEFTDSRGSKMT--SDGINDKK-VYMLPGGDSLNGPHNPINETGLV-L- 384
Db      287  ERPHFVFOGDKXYLFTISH--KYTFADNLTGPDGYV--GFVSDPKLTGPYTPMNSSGLVIG 342
Qy      385  MNMLNDPADLTHYSHCGIFHPGCGNNVVLTSYM 416
Db      343  NPSSOP---FOYTSYHVP-----NGLVYSFI 366

```

Result	12			
SACB_ZYMMO		STANDARD;	PRT;	423 AA.
ID	SACB_ZYMMO	Q60114; Q60116; Q06487;		
AC	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Levanucrase (EC 2.4.1.10) (beta-D-fructofuranosyl transferase)			
DE	(Sucrose 6-fructosyl transferase).			
GN	SACB OR LEVU OR SUCE2.			
OS	Zymomonas mobilis.			
OC	Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;			
OC	Zymomonas.			
OX	NCBI_TaxID=542;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 10988 / ZML;			
RX	MEDLINE=93305726; PubMed=8318541;			
RA	Song K.B., Joo H.K., Rhoe S.K.;			
RT	"Nucleotide sequence of Levanucrase gene (levu) of Zymomonas mobilis			
RL	ZML (ATCC10988).";			
RL	Biochim. Biophys. Acta 1173:320-324(1993).			
RN	(2)			
RP	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.			
RC	STRAIN=IFO 13756;			
RX	MEDLINE=95218269; PubMed=7766026;			
RA	Kyono K., Yanase H., Ionomura K., Kawasaki H., Sekai T.;			
RT	"Cloning and characterization of Zymomonas mobilis genes encoding			
RL	extracellular levanucrase and invertase.";			
RL	Biotech. Biotechnol. Biochem. 59:289-293(1995).			
RN	(3)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NRRL B806;			
RA	Gunasakthan P., Mukundan G., Kannan R., Velmurugan S.,			
RA	Alt-Adelkader N., Alvarez E., Baratti J.;			
RL	Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.			
RN	(4)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 31821 / ZM4 / CP4;			
RA	Ahn J.Y., Kang H.S.;			
RT	"Sequence analysis of 44B6 fsmid clone of Zymomonas mobilis ZM4.";			
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.			

```

CC      -I -CANAATIC ACTIVITY: Sucrose + ((2,6)-beta-D-fucosyl21)(N) =
CC      glucose + ((2,6)-beta-D-fucosyl24)(N+1).
CC      -I -SUBCELLULAR LOCATION: Secreted.
CC      -I -PM: DOES NOT SEEM TO BE N-TERMINALLY PROCESSED.
CC      -I -SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF081588; AAA27695.1; -.
DR      EMBL; D17524; BAA04475.1; -.
DR      EMBL; L33402; AAA27702.1; -.
DR      EMBL; AF313764; AAG29870.1; -.
DR      InterPro; IPR003469; Glyco_hydro_68.
DR      Pfam; PF02435; Glyco_hydro_68; 1.
DR      KW      Transferase; Glycosyltransferase.
FT      CONFLICT 39          V -> I (IN REF. 2).
FT      CONFLICT 200        NPED -> TPXI (IN REF. 1 AND 4).
FT      CONFLICT 217        E -> Q (IN REF. 2).
FT      CONFLICT 220        T -> A (IN REF. 2).
FT      CONFLICT 244        C -> Y (IN REF. 2).
FT      CONFLICT 379        I -> V (IN REF. 2).
SQ      SEQUENCE 423 AA; 46762 MW; 19A4691DA3BB9FE3 CRC64;

```

Query Match	9.9%: Score 243; DB 1; Length 423;
Best Local Similarity	24.4%: Pred. No. 1,1e-08;
Matches	108; Conservative 60; Mismatches 135; Indels 140; Gaps 25;
OY	16 TRADMLKIPGQNSPQKVFQPFNASAIKNIDSAKGYKSGNLDLDL-----VWDSW 66
DB	14 TRADAMKY--HTDPTATMP-----TIDYDPVMTKYYWVDTW 50
OY	67 PLQADGTAANYHGYHIVSALAGD-----PKNSDPTPLHLFYQKGDTSIDSKNNAGRV 120
DB	51 PLRQINGVVSFOQMSVIFALVADRTKYGCMNRNDGARIGIFYSRGG-----SNWIFPGHL 106
OY	121 FEDMDKVPNDPYLYKOTQEWSSGATLTKDG--QVRLEFTDYSGNPEDGDTGAGNOIIS 177
DB	107 LKD-----GAMP-----BSWMSG--CTIMAPGTANSVEFFFSVNDTP----- 143
OY	178 TAQVNLSPDPAATLKVQDVSCHKSV-EPG-----GDCTVQNIQOQFIDECKWISGD 227
DB	144 -----SESVAQCKGKGIYADDKRSWMDGPKFYATDLQADGLIYAD---YAENNFW--- 190
OY	228 NHLRLDPHYV--EDKGHKLYLFEANTGTTDYGQDGSFNKAKYGGSDVFPQNEKNKLLQ 285
DB	191 --DPRDPHFVETNPBDGKTYALFEENV-----AMERGTVAAVEEGEIGPY- 231
OY	286 SPKROI--ASLANGALGIYE-LADDYTVASVMKPLVASMTVDADEVSRANIFKNNKAYL 341
DB	232 PPKETEPDGAATYCAAIGIAQALAEARTEMKLLRPPLPTATAGVNDOTRPHVYPONGLTYL 291
OY	342 FTDSRSGKMTSDGINDKDYVYMLGFG-----DSLNGPHNPINETSGLYL-NMNLDPADL 393
DB	292 FTIHSHTYV-ADGLS-----GPDGYGVGFSENGICFGRPEPLNGSLVYLGPNSSQP--- 340
OY	394 THHTYSHGCIHPREGNNVYLTSM 416
DB	341 YQAYSHYV-----TNGLVTSFI 358

FT	CONFLICT	1698	1698	D -> N (IN REF. 1)
FT	CONFLICT	1795	1798	SA -> T (IN REF. 1).
FT	CONFLICT	1836	1837	OY -> T (IN REF. 1).
SO	SEQUENCE	1953	AA; 200150 MW; 611B3FLC9549D91AE CRC64;	
Query March				
Best local similarity 5.7%; Score 139; DB 1; Length 1953;				
Matches 109; Conservative 59; Mismatches 199; Indels 142; Gaps 23;				
QY	54	SGNLIDLDVW-----DSWPLONADGTAANYHGVIYSALAGDPKNSDDPRLHLFLYOKVG	107	
Db	278	NGSVLEGVVWVYSHQDNQWOLTTADGKTLANTGMDVDYDANAAYIEGTEGNELLYKKYDSRG	337	
QY	108	DTSIDSMKNAGRVEEDMKREVPNDPYLKQYTOEMSGSA-----TLTKDQVRLFYTD	159	
Db	338	YLIIAD--DMTIVYISGDQAHSNRGMDISGDRGTGVIISGDRVYVNLTGSSV-----TD	391	
QY	160	YSGNPEEGCGTCAGQIIISIAQVY-----LSQPDAAAT	190	
Db	392	GATGMVVISGDDTNTMTISGHSIVDANATGALISGNETTTFACDIAVSGGTAIIIDGNAT	451	
QY	191	LKVDGVDHDK-----SVFDGCGDTGVYQNIQOFIDE--GKVISGDNHTLDRPH-----	235	
Db	452	IKNIGTSDISAGSTGYIDGNNAKRVNNDGIMTITDGTGSHITGDVNVIDNAGSTTVSG	511	
QY	236	-----YVEDKGKRYLF-EANTGTTCGYOGDOSFNKKAY-----YGSDFEFQ	277	
Db	512	ADATAYLIE--GDNALVINEGNIITSGAVGTRIDGDDAHTNTGDIADVAGSAVAIIN	569	
QY	278	NEKKKLLQSPKKQJASLANCALGIVELADDTYKSVKPLVASNTVADEV-----ERANI	332	
Db	570	GDNGSLTQAGGLV---TDGAMGIITVCTGMEANNTGNATVRD--ADSVGFVVAEGKNT	623	
QY	333	FKANNKMYLFTDSRG-----SKMTSDG-----INDKDYVMLGPGGSLNGPHNP	377	
Db	624	FK--NKGDIDVSLNGTALVSGDMSQVTLDDGIDNVYSQDSEGVFSATGVSVSGDSNAV	681	
QY	378	NETGLVINMLNDPADLTHTVSHCIP-----HPEGNNVYLTSM-----TNRGFY	422	
Db	682	DITG---NVNIS-ADYGDDDLAAGAPPLTGVVVGNGNCTVTLNGALNTIDDDLSATGQY	737	
QY	423	PEHSHLRDKLGAVINKSDT-----SGCEN	447	
Db	738	-----LDVVGLSVTGGDDNDVEIDGGIN	759	
RESULT 14				
GTFL_STRDO	GTFL_STRDO	STANDARD;	PRT: 1597	AA.
ID	AC	P11001;		
DT	01-JUL-1989	(Rel. 11, Created)		
DT	01-JUL-1989	(Rel. 11, Last sequence update)		
DT	15-DEC-1998	(Rel. 37, Last annotation update)		
DE	Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)			
DE	(Sucrose 6-glucosyltransferase).			
GN	GTFI.			
OS	Streptococcus downei (Streptococcus sobrinus).			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_Taxid=1317;			
RN	11			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MEP28;			
EX	MEDLINE=87308014; Pubmed=3040686;			
RA	Ferretti J.J., Gilpin M.L., Russell R.R.B.;			
RT	"Nucleotide sequence of a glucosyltransferase gene from Streptococcus			
RT	sobrinus MFE28."			
RL	J. Bacteriol. 169:4271-4278(1987).			
CC	-I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS THAT ARE THOUGHT			
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE			
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE			
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.			
CC	-I- CATALYTIC ACTIVITY: Sucrose + [(1,6)-alpha-D-glucosyl](N) = D-			

```

CC fructose + ((1,6)-alpha-D-glucosyl)(N+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M17391; AAC63063.1; -
CC InterPro: IPR002479; CW_binding.
CC InterPro: IPR003318; GH_70.
CC Pfam: PF01473; CW_binding_1; 19.
CC Pfam: PF03324; Glyco_hydro_70; 1.
CC Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
CC SIGNAL 38
CC CHAIN 1 1597
CC FT DOMAIN 39 1597 CATALYTIC (APPROXIMATE).
CC FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).
CC FT DOMAIN 1099 1597
CC FT REPEAT 1099 1132 1.25 A, 2 B, AND 5 AC REPEATS.
CC FT REPEAT 1163 1213 A REPEAT.
CC FT REPEAT 1227 1277 AC REPEAT.
CC FT REPEAT 1292 1342 AC REPEAT.
CC FT REPEAT 1352 1399 B REPEAT.
CC FT REPEAT 1406 1455 AC REPEAT.
CC FT REPEAT 1465 1512 B REPEAT.
CC FT REPEAT 1519 1568 AC REPEAT.
CC FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).
CC FT SEQUENCE 1597 AA; 177080 MW; B9E86A20868798E CRC64;
SQ
Query Match 5.5%; Score 135; DB 1; Length 1597;
Best Local Similarity 22.9%; Pred. No. 0.53;
Matches 80; Conservative 30; Mismatches 111; Indels 128; Gaps 19;
QY 36 QPNASAIKNIDSAGYDKSNMLIDLVDSWPLONADG-----TAANYGHYHISA-- 86
DB 1209 QHNQNAVNTNVA---DKTGH-----WYLGKDSVAVTGAQTYGKQHLFEANGQ 1256
QY 87 LAGDPKNSDPTPLHLFYOKVGDTSIDSWKN-----AGRVFEDMKREVPNDP----- 132
DB 1257 VKGDFVFAKDKGLFYVDVDSG-----DMWTNFIEDKAGNMF-----YLGKGAAVTGAQT 1307
QY 133 -----YLYKQTOEMSGSATITLKDGQVRLFTDYSGNEDDGTGAGNIIITAVYNLSQP 186
DB 1308 IKGKGLFKFANGQOQKGVIVKDAQCKIRYV-----DAQT--GEQVEN----- 1347
QY 187 DAATLKVDGVS DHKSVFEDGSGDTYVQNI-----QOQFID-----EGKMTSGDNHTLR 232
DB 1348 ---KSVVNG---KTYFGSGDTAQTQANPKGQTFKSSGVLRFPYNLEGOVSSGWET 1401
QY 233 DPH-----YVEDKGH--KYLVEFANTGTT---DGYQDQSF-- 263
DB 1402 AEHEMVVYVSKGVLTAQTIGNQRVYFKDNGHOVKGGLVGTGNDGKLYRYDANSGDQAFNK 1461
QY 264 -----NNKAYVGGSDVFPQNEKNKLLQSPKQIADSLANALGIVELADY 308
DB 1462 SVTVNGKTYTFHFGSDGTAQTQAN-----PKGQTFKDSGVLRFPYNLEGOY 1505

```

```

DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN GTFB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OC NCBI_TaxID=1309;
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239, MT4245, MT4251, MT4467, and MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) = D-
CC fructose + ((1,6)-alpha-D-glucosyl)(N+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M17361; AAA88588.1; -
CC EMBL; D88651; BAA26101.1; -
CC EMBL; D88654; BAA26105.1; -
CC EMBL; D88657; BAA26109.1; -
CC EMBL; D88660; BAA26113.1; -
CC EMBL; D89977; BAA26119.1; -
CC PIR; B33135; B33135.
CC InterPro: IPR002479; CW_binding.
CC InterPro: IPR003318; GH_70.
CC Pfam: PF01473; CW_binding_1; 13.
CC Pfam: PF03324; Glyco_hydro_70; 1.
CC Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
CC SIGNAL 34
CC CHAIN 1 1476
CC FT DOMAIN 35 1476 GLUCOSYLTRANSFERASE-I.
CC FT DOMAIN 1097 1476 CATALYTIC (APPROXIMATE).
CC FT DOMAIN 1097 1130 GLUCAN-BINDING (APPROXIMATE).
CC FT REPEAT 1097 1130 A REPEAT.
CC FT REPEAT 1161 1210 5 X TANDEM REPEATS.
CC FT REPEAT 1225 1275 1.
CC FT REPEAT 1290 1340 2.
CC FT REPEAT 1355 1405 3.
CC FT REPEAT 1420 1470 4.
CC FT REPEAT 1420 1470 5.
CC VARIANT 62
S -> T (IN STRAIN MT4239).

```

Db 1432 -----GARTINGOH 1440
Search completed: June 16, 2003, 12:07:27
Job time : 30.9037 secs

```
FT VARIANT 65 65 I -> T (IN STRAINS MT4239, MT4245,  
FT VARIANT 68 68 MT4251, MT467 AND MT8148).  
FT VARIANT 78 78 A -> V (IN STRAIN MT4239).  
FT VARIANT 86 86 O -> P (IN STRAIN MT4251).  
FT VARIANT 89 89 S -> I (IN STRAIN MT4239).  
FT VARIANT 168 168 S -> F (IN STRAIN MT4251).  
FT VARIANT 168 168 K -> N (IN STRAIN MT4251).  
FT VARIANT 276 276 D -> S (IN STRAINS MT4239, MT4245 AND  
MT4251).  
FT VARIANT 399 399 N -> R (IN STRAIN MT4239).  
FT VARIANT 474 474 I -> T (IN STRAIN MT4239).  
FT VARIANT 512 512 K -> R (IN STRAIN MT8148).  
FT VARIANT 519 519 F -> Y (IN STRAIN MT8148).  
FT VARIANT 701 701 T -> I (IN STRAIN MT8148).  
FT VARIANT 708 708 A -> V (IN STRAIN MT8148).  
FT VARIANT 938 938 F -> L (IN STRAIN MT8148).  
FT VARIANT 952 952 YGTPVA -> FGKPE (IN STRAINS MT4245,  
MT4251 AND MT8148).  
FT VARIANT 963 964 MT -> SV (IN STRAINS MT4245, MT4251 AND  
MT8148).  
FT VARIANT 968 970 VDG -> ADS (IN STRAINS MT4245, MT4251 AND  
MT8148).  
FT VARIANT 1086 1086 A -> T (IN STRAIN MT4239).  
FT VARIANT 1158 1158 S -> N (IN STRAIN MT4239).  
FT VARIANT 1163 1163 H -> Y (IN STRAIN MT4251).  
FT VARIANT 1168 1168 E -> K (IN STRAIN MT8148).  
FT VARIANT 1182 1182 Y -> C (IN STRAIN MT8148).  
FT VARIANT 1234 1234 A -> P (IN STRAIN MT4239).  
FT VARIANT 1263 1263 H -> P (IN STRAIN MT8148).  
FT VARIANT 1263 1263 H -> R (IN STRAINS MT4239, MT4245 AND  
MT4251).  
FT VARIANT 1264 1264 H -> Y (IN STRAINS MT4245 AND MT4251).  
FT VARIANT 1272 1272 G -> S (IN STRAINS MT4245 AND MT4251).  
FT VARIANT 1329 1329 Y -> H (IN STRAINS MT4239, MT4245, MT4251  
AND MT8148).  
FT VARIANT 1394 1394 H -> Y (IN STRAINS MT4245 AND MT4251).  
FT VARIANT 1402 1402 G -> S (IN STRAINS MT4245 AND MT4251).  
FT VARIANT 1459 1459 Y -> H (IN STRAIN MT4467).  
FT CONFLICT 570 570 R -> A (IN REF. 1).  
FT CONFLICT 800 817 ADOOVRAASAPSTDGK -> LIKMFALRLRPHQQA  
(IN REF. 1).  
FT CONFLICT 1310 1310 H -> L (IN REF. 1).  
SQ SEQUENCE 1476 AA: 165685 MW: 3479B62B07694D98 CRC64;  
  
Query Match 5.5%; Score 134; DB 1; Length 1476;  
Best Local Similarity 21.3%; Pred. No. 0.56;  
Matches 80; Conservative 47; Mismatches 113; Indels 135; Gaps 21;  
  
QY 68 LQNAAGTAANYHG-----YHIVSALAGDPKNSDDTPLHLFYOKVGPDSIDSWKNA 117  
Db 1132 LKNEDGTYYA--YGGNDGRRENGYQFMGVRHFNNGE-----MSVGLTVID----- 1177  
QY 118 GRV--FEDMDKRVNDRLKYQTOEMSGSATLTDCGOVRLFTYDYSN----- 163  
Db 1178 GQVOYFDEMG-----YQAK--GKEVTTADGKIR-YFDKQSGNMVRRRIENEDEG 1223  
QY 164 -----PEDGGTAGNOIISTAOVNLSDPAATLKVDSGDHK--SVFDGGDGTGYONIQ 215  
Db 1224 KMLYLGEDGAALVYGSQITN-GQHLXFRANGVQVKGFEVTDHGRISTYDGSNGDOIRN-- 1280  
QY 216 QFT--DEGKWTSGDNH-----TLRDPHY-----VEDGKHRYLVEEANTGTTDGYOG 259  
Db 1281 RFVFNAGQWFEYFDNNNGYAVTGARTINGOHLYFRANGVQVKGFEVTDRYGRISTYDGSNG 1340  
QY 260 DQ-----SFNNKAY-----YGSDFEFOENKKNLLOSPKCIASLANGA 298  
Db 1341 DOJRNRFVRNAQGWFEYDNNNGYAVTGARTINGOHLYFR-----ANG- 1382  
QY 299 LGIYELADDTYKSVKPLVASNTVADEVERANIFKMNKNKYLFPTDSRGSKMTSDGINDK 358  
Db 1383 ---VQVAGGEVTDHGRISTYDGSNGDOIRNRFVRNAQGWFEYF-DNNGYAVT----- 1431  
QY 359 DVTWMLGPGGSLNGPH 374
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 12:01:28 ; Search time 114.682 Seconds
(without alignments)
817.488 Million cell updates/sec

Title: US-09-986-682B-3
Perfect score: 2454
Sequence: 1 MNSDYEKEDYGFHITRADM.....NIKSDTSGGENSGCGQFP 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2454	100.0	487	2	082854
2	1752	71.4	489	16	097181
3	1443	58.8	499	2	092553
4	300	12.2	415	2	09AEX9
5	295	12.0	428	16	097179
6	291	11.9	630	2	09EVD6
7	283	11.5	431	2	093TM1
8	278.5	11.3	424	2	093FV9
9	278.5	11.3	578	2	08W87
10	240.5	9.8	430	2	09LBN1
11	141	5.7	1141	16	0932F7
12	141	5.7	1449	2	09F3Z5
13	141	5.7	1449	16	09JWB4
14	139	5.7	1141	16	099W46
15	138	5.6	1166	2	086485
16	133	5.4	993	16	08Y105

17	132.5	5.4	1377	2	088093	088093 escherichia
18	132.5	5.4	1377	2	047692	047692 escherichia
19	129	5.3	728	16	08ZHR2	08ZHR2 yersinia pe
20	127	5.2	1983	16	0931N9	0931N9 staphylococ
21	127	5.2	2186	16	099TBO	099TBO staphylococ
22	126	5.1	13055	5	009165	009165 caenorhabdi
23	125.5	5.1	1577	2	054178	054178 streptococc
24	125	5.1	759	16	082CG3	082CG3 yersinia pe
25	125	5.1	2015	16	09JRD2	09JRD2 neisseria m
26	125	5.1	2110	2	09LBN1	09LBN1 pasteurella
27	122	5.0	809	5	P90534	P90534 dictyostell
28	122	5.0	1136	2	09X686	09X686 salmonella
29	121	4.9	555	17	029748	029748 archaeoglob
30	120.5	4.9	3262	16	08YTO8	08YTO8 anabena sp
31	120	4.9	819	2	09X583	09X583 streptomyce
32	120	4.9	1849	2	09S4K2	09S4K2 lactobacill
33	120	4.9	2119	2	09AH75	09AH75 streptococc
34	120	4.9	2140	16	097RY6	097RY6 streptococc
35	120	4.9	2144	2	09S4M8	09S4M8 streptococc
36	119.5	4.9	519	12	09OB31	09OB31 human adeno
37	119.5	4.9	519	12	096618	096618 human adeno
38	119.5	4.9	1123	2	09ZIB3	09ZIB3 wolineella r
39	119.5	4.9	2523	17	08TJS8	08TJS8 methanosarc
40	119	4.8	1198	17	08TP97	08TP97 methanosarc
41	118.5	4.8	1075	17	08TP22	08TP22 methanosarc
42	118	4.8	1727	4	09UP09	09UP09 homo sapien
43	118	4.8	1755	17	026812	026812 methanobact
44	118	4.8	4163	2	09LAE6	09LAE6 rhizobium l
45	117.5	4.8	263	2	09AN41	09AN41 bradyrhizob

ALIGNMENTS

RESULT 1

ID 082854 PRELIMINARY: PRT; 487 AA.
AC 082854;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Beta-fructofuranosidase.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V230;
RA Tsusaki K.;
RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=V230;
RA Kurimoto M.;
RT "Cloning and sequencing of beta-fructofuranosidase gene from Bacillus
RT sp. V230."
RL Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB010272; BAA32083.1;
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
SQ SEQUENCE 487 AA; 53412 MW; 7F52A47921824AE3 CRC64;

Query Match 100.0%; Score 2454; DB 2; Length 487;
Best Local Similarity 100.0%; Pred. No. 1.5e-146;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSGDYKEDYGFHITRADMLKIPGOONSPOKVFQFNASAIKINDSAKGYKSGNLIDL 60
|||||
DB 33 MNSGDYKEDYGFHITRADMLKIPGOONSPOKVFQFNASAIKINDSAKGYKSGNLIDL 92
|||||
QY 61 DWSDWPLQADGTAANYHGHTVSALAGDPKNSDDTPLHLFYQVGGTSDISWNAAGY 120
|||||

Db 327 IVELNDYTLKNVMPPLIASLVYDEIERANVFKMGLMVLFTSGSKYTVDAIGDDI 386
OY 361 YMLPGGDSLNGPNHINETGLVLMNLDPADLTHYSHGCIPEEGNNVLTSTYMTNRG 420
Db 387 YMLGIVSTSLTGPKPLNGTGLVHODLDRDVTWTAHFAIPGKNNVVSSTYMTNRG 446
OY 421 FYPEHSHLRDLKGLVNTKSGDSTGSGNS-SCGQ 453
Db 447 LEPDKSTFAPSFLNKGSKTSYVKNGLIEGQ 480

RESULT 4

O9AEX9

ID O9AEX9

AC O9AEX9

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)

DE Levensaustrase.

OS Rahmella aquatilis.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Rahmella.

OX NCBI_TaxID=34038;

RN [1]

RP SEQUENCE FROM N.A.

RA Kim H., Cha J.;

RT "Cloning and characterization of levanucrase";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY027657; AAK14794.1;

DR InterPro; IPR003469; Glyco_hydro_68.

DR Pfam; PF02435; Glyco_hydro_68; 1.

SQ SEQUENCE 415 AA; 45958 MW; F09F5A1BF8690C5E CRC64;

Query Match

Best Local Similarity 26.7%; Score 300; DB 2; Length 415;

Matches 116; Conservative 60; Mismatches 150; Indels 108; Gaps 22;

OY 16 TRADMLKIPGQNSPOKRVPOFNASAIKNIDSAGYKSGNLDIDVDSWPLQADGTA 75
Db 12 TRADALV--NENPTTQPIVSADFPYMSD-----EYFIMDAFPLSLDSTV 57
OY 76 ANYGHIYVSLADGPKNSDDTPLH-----LEFYQVGDTSIDSW 114
Db 58 VSVGWSYIETLTAQRNNNNSEYLDADGNYDTSDMNNRHRGAKICWYSRTG---KDW 113
OY 115 KNAARVEDMDKFPNDPYLKQTOEWSGS-ATLKQGVPLFTYDYSNGNPDGSGAGN 173
Db 114 IFGRVMAE-----GVSPTREWAGTPILLNEDEVDLYTTCV-----PGA 155
OY 174 QIISTAQVNLSPDAATLKVDGVSDFDGDGTVQNTQOFIDEKWTISGDNHTLRD 233
Db 156 TIAVRGKVLTSSEGVTLA--GNEVKSLEF-SADGVYIQTESQ---NTYNNFRDPSPIID 209
OY 234 PHYEDKGHYLVEANTGTTDYGQDQSFNNKAYYGGSDVFQNEKNKLLQSPKKQIAS 293
Db 210 PH---DGKLYMVEGVNA-----GER-----GSHVIGKQEMGTL--PRDHRDVG 248
OY 294 LANGALGIVELA-----DDYTVKSVMKPLVANSYVADEVERANIEFMNKKVILFTSR 346
Db 249 NARQACIGMAVAKDLISGDEM---EILPLVTAAGVNDQREPERHEFQDCKYLLFTTISH 305
OY 347 GSKMT-SDGINDKD-VYMLPGGDSLNGPNHINETGLVLMNLDPADLTHYSHGCI 404
Db 306 -KFTYAEGLGPGGVY--GFLSDRLTGPGSPMNGSLVGLGN--PPSQPQIYSHCWNP- 358
OY 405 PEGNNVLTSTYMTN 418
Db 359 ----NGLVTSEFIDN 368

RESULT 5

O97179

ID O97179

PRELIMINARY;

PRT; 428 AA.

OY

DB

DT

DE

OS

OC

OX

AC O97179;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Levensaustrase.
GN CAC1774.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hill J., Wolf Y.I.,
RA Tatusov R.L., Sabatne F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007686; AAK79739.1.
DR InterPro; IPR003469; Glyco_hydro_68.
DR Pfam; PF02435; Glyco_hydro_68; 1.
DR Complete proteome.
SQ SEQUENCE 428 AA; 48395 MW; 9670B154B178E23E CRC64;

Query Match

Best Local Similarity 27.7%; Score 295; DB 16; Length 428;

Matches 112; Conservative 62; Mismatches 147; Indels 84; Gaps 21;

OY 59 DLDVDSWPLQADGTAAANYGHVLSAL-----AGDKNSDDPPLHLYQKVGDTSDS 113
Db 66 NLWVMDTPPLVKKDGLAVNMGKYIFALTSRNVGMKRRHDVAGISYFCSTDG-----EN 121
OY 114 WKNARVEEDMDKFPNDPYLKQTOEWSGATLKQGVPLFTYDYSNGNPDGSGAGN 173
Db 122 WYKGLAVNVEDAL-----GSRQWAGSAILDENGMQFEFT-----ATGRKG 163
OY 174 QIISTAQVNLSP-----DAATLKVDGVSDFDGDGTVQNTQOFIDEKWTISGDNHTLRD 227
Db 164 EAVRTEQRLVTKFINSIVNDKGVHITNCSKHQVILE-PDGVYVYQTPMOO--AKGPII--- 217
OY 228 NHTLMDPHYVEDPKGHK--YLVEANTGTTDYGQDQSFNNKAYYGGSDVFQNEKNKLLQ 285
Db 218 -YSFEDPYFFEDPKTKKYLIFEGNKGGKIEKMPENIGDKLF-----RNHILAP 266
OY 286 SPKKQIASLANGALCI-VELADYTVKSVMKPLVANSYVADEVERANIEFMNKKVILFTD 344
Db 267 RGVENF-----NGNVGIAVAOKKDLRFKLLPLLEAVGVNQLERPQIVMKKNKYLLFTI 322
OY 345 SRGSKMT-SDGINDKOVYMLPGGDSLNGPNHINETGLV--NMNLDPADLTHYSHGCI 402
Db 323 SH--KFTYAOGLNGVD-GLYGFCGNSLRSNKKPLNGNLVITNPIND---YQTSWYLV 376
OY 403 PHPEGNVNLVSYMTNRGFEPEH--SHLR-----DKLGVIK 439
Db 377 ---SGHDVL-----SFINEYHFNGLRGCTGTAPLTQISLKG 410

RESULT 6

O9EVD6

ID O9EVD6

AC O9EVD6

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Fructosyltransferase.

GN FTF

OS Actinomyces naeslundii.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Actinomycetaceae; Actinomycetaceae; Actinomycetaceae;

OX NCBI_TaxID=1655;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 12104.
 RA Bergeron L.J., Moron-Bermudez E., Burne R.A.;
 RT "Characterization of the fructosyltransferase gene of *Actinomyces*
 naesiundii";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF228582; AAC09737.1; -
 DR InterPro: IPR003469; Glyco_hydro_68.
 DR Pfam: PF02435; Glyco_hydro_68; 1.
 KW Transferase.
 SQ SEQUENCE 630 AA; 68256 MW; CCA04184DD6D7D2 CRC64;

 Query Match 11.9%; Score 291; DB 2; Length 630;
 Best Local Similarity 25.3%; Pred. No. 2.6e-10;
 Matches 133; Conservative 72; Mismatches 208; Indels 112; Gaps 29;

 QY 7 KEDGFHATIRADMIKT-----PQONSPOFKVPOFNASAIKNIDSAGYDKSGNLIDL 60
 DB 120 EENY-TAKMTRADAMQIOIRVFQPOGAVRAPRLPEQLMPEISN-----GEPATSE--DV 171
 QY 61 DWDSMPLOMDGTANHGHIYSALAGDPK---NSDDTPLH---LFYQKV----- 107
 DB 172 WWDWTWLTDEAHOISYNGWEIAFSLVADRHAGYTFEDRPHARLGFYRKAGTQTSSA 231
 QY 108 ---DTSIDSMKNAGHFEEDMDKFEVNDPYLKYOTQEMSGSATLTKDGOVRLFTYDYGNP 164
 DB 232 DGAGSSNGGWYIGGHVFPDAS--ASSRTYHAQTEMGSGARLMGKNIRHFTYSVSTT 289
 QY 165 EDGGTG-AGNOIISTAQVNLSP-DAATLKYDGVSDHKSVPDGGDGYVONI---OOFI 218
 DB 290 APPRGTTHQVIAHT---SKPYDPRIVQSEG---RIYATKNGVLTGFTFTHQLLV 340
 QY 219 DEGWKIS---GDNHTLRDPHYVEDKGH-----KYLEFANNGTTDYGOGDSFNKA 267
 DB 341 PDGKTYOTREONPGVN--FRDPFTFRQNNPSDPEIWEVFGNS---AFVREQIYDAA 394
 QY 268 YYGSGDVEFFON-----EKNKLQSPKKQIASLANGALGIELAD-----DYTV 310
 DB 395 AKAGONTLATCTEEDLEGYEGK---DPKAEIVAVNQRGVYQDLANGLARAKNKAMTE 450
 QY 311 KSVMPVLVASTVADVEVRANIEFKNNKNWYLFTRSGSKMTSDGINDK-VYMLGPGGDS 369
 DB 451 WEYLPRLISGNCVNDQTERPQIYHODGKYVLFITISH-REYVADIGLOGEGVY--GPGDG 507
 QY 370 LNPENPILNE-TGLVLT---NMNLDPAD-----LFTYSHGCIHPPEGNNV--V 411
 DB 508 LRSQKPLNONTGIALGNPILNPNPKPKYSPDENQSPRTYQSTSHYMP---GGLVESF 564
 QY 412 LSTYTNRGFYPEHHSHLRDLKLVNIKSGDTSGE--NSSGQGF 454
 DB 565 IDSIGNKDGNPVRGSGSLPTVKLINISGDTTSVDRYCTGNLGLGF 609

 RESULT 7
 Q93TMI PRELIMINARY; PRT; 431 AA.
 AC Q93TMI;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Levanusucrase LscC (EC 2.4.1.10).
 GN LSCC.
 OS *Pseudomonas syringae* (pv. *glycinae*).
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=318;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PG4180;
 RX MEDLINE=98394981; Pubmed=9726857;
 RA Heltner U., Jaechel F.R., Boch J., Meyer M., Rudolph K., Ullrich M.S.;
 RT "Cloning, nucleotide sequence, and expression in *Escherichia coli* of

RT levanusucrase genes from the plant pathogens *Pseudomonas syringae* pv.
 RT *glycinae* and *P. syringae* pv. *phaseolicola*.";
 RL Appl. Environ. Microbiol. 64:3180-3187(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PG4180;
 RX MEDLINE=21242726; Pubmed=11344135;
 RA Li H., Ullrich M.S.;
 RT "Characterization and Mutational Analysis of Three Allelic lsc Genes
 RT Encoding Levanusucrase in *Pseudomonas syringae*.";
 RL J. Bacteriol. 183:3282-3292(2001).
 DR EMBL; AF346402; AAK49952.1; -
 DR InterPro: IPR003469; Glyco_hydro_68.
 DR Pfam: PF02435; Glyco_hydro_68; 1.
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 431 AA; 47619 MW; 0C71D0F133071AA5 CRC64;

 Query Match 11.5%; Score 283; DB 2; Length 431;
 Best Local Similarity 26.2%; Pred. No. 4.8e-10;
 Matches 116; Conservative 63; Mismatches 135; Indels 128; Gaps 25;

 QY 16 TRADMKTIGQONSPOFKVPOFNASAIKNIDSAGYDKSGNLIDLDWDSMPLOMDGTA 75
 DB 28 SRADALKV--NENDPTTQPLVSADFPVNSDT-----VFIMDTMPLRELDTGV 73
 QY 76 ANYGHIYSALAGDPKNSDDTPLH-----FYQVGDTSID 112
 DB 74 VSNAGSVILTLAD-RHEDD-PQYLDANGRIDKRDWEDRGRARMCYWSRTG-----K 127
 QY 113 SMKNAGFEEDMDKFEVNDPYLKYOTQEMSGSATLTKD-GQVRLFTYDYGNEPDDGTA 171
 DB 128 DWIFGCAVMAE-----GVSPTREWAGTPIILNDGIDLIVYTCCT----- 168
 QY 172 GNOIISTAQVNLSPDAATLKYDG---VSDH-----KSYPDGGGTYVYONI 218
 DB 169 -----PAAAVAKVGRIVISDQGVELKDTQVKKLE-ADGTYVTEAQ-- 211
 QY 219 DEGWKISGDNHTLRDPHYVEDKGHLYFEAVNTGTTDYGOGDSFNKNAYGSDVFFON 278
 DB 212 -NSSMNPDPSPETIDP-----NDGKLYMVEGVNA---GGRGSH-----VGAALGVP 257
 QY 279 EKKKLQSPKKQIASLANGALGIELADDTYVK--SVKPLVASTVADVEVRANIEFKN 336
 DB 258 PGHEIDVGARFQY---GCIGLA-VAKDLSEEMELIPPLVTAAGVNDQTERPHYVPOD 311
 QY 337 NKMYLFTDSRSGKMT--SDGINDK-VYMLGPGGDSLNGPHNFINFGVLNMLNDPADLT 394
 DB 312 GKTYLFTISH--KFTYADGITGPDGY--GVGEHLFGFIRPMNASGLVLR--PPEOPF 365
 QY 395 HTYSHGCIHPPEGNNVLTYSY 416
 DB 366 QFYSHCVMP-----NGLVTSFI 382

 RESULT 8
 Q93FUG PRELIMINARY; PRT; 424 AA.
 AC Q93FUG;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Levanusucrase.
 GN LSCA.
 OS *Pseudomonas aurantifaci*.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=86192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S-4380;
 RA Koh I., Jang E.K., Kim J.S., Song K.S., Kim C.H., Rhee S.K.;
 RT "Screening, molecular cloning and characterization of a novel
 levanusucrase, and its expression in *Escherichia coli*.";

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF306513; AAL09386.1; -
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1
DR SEQUENCE 424 AA; 47009 MW; 9C1B4936A985162E CRC64;

Query Match 11.3%; Score 278.5; DB 2; Length 424;
Best Local Similarity 24.7%; Pred. No. 9,1e-10;
Matches 118; Conservative 70; Mismatches 147; Indels 143; Gaps 27;

16 TRADMLKIPGOONSPOFVPOPNAS--AIKNDIAKGYDKSGNLIDLVWDSPLQNDG 73
19 TRADMLK--HADPTTQPLVSADFPVLSN-----EVEFWTMTPLRDLG 62
74 TAANYHGHIVSALA-----GDPKNSDD-----TPLLFYOKVYDTSID 112
63 NITSVDGSHVFTLTADHPNDPEYIDENGNYDITRDMNDRHRAKMTYWSRTS----K 118
113 SWKNAGRVFEDMDKFPVNDPLKTYOTOMSGS-ATLTGQGVRLFTYDYSNPEDEGTGA 171
119 DMKLGGRVMAE-----GVSPTAREMAGPILLNCGDVLTYTAVT----- 159
172 GNOIISTAQVNLSDPDAATLVDC---VSDH-----KSVFDGDCGYVYQNIQFI 218
160 -----FGATIVYRGVRYVTEHGVSMVGEKVPLE-ADGKMYQTEAQ-- 202
219 DEGWISGDNHTLRDPHYVEDEKGHKYLVEANVTGTTDYGQDGFNNKAYVGGSDVPFN 278
203 -NPEMGFDPMPFPRDP-----NDGKLYLMEGVNA---GERGSHKV-GKAELGIDVPPGTE 253
279 EKNKLLQSPKQOIASLANGALGIY---ELADDTYKVSVMKPLVANSNTVADEVERANIEK 334
254 VGNSEFGQ-----ACVGIAVARDEDDDM---EMLPPLLAAGVINDQTEPHVYF 300
335 MNNKWTLETDGSGKMTSDGINDKD-VYMLGPGGSLNGPNNPIETGLVL-INMLDPA 392
301 QDGKYLETTISH-FTYVADGVTPGVY--GFVADSLGCPVPLNGSLVLCNPSSQ-- 355
393 LHTYSHGCIHPHEGNNVLTYSMTNRGFYEPHSHLDKLG-----VNIKSDT 442
356 -FQYTSCHVMP-----NGLVTSFIDS---VPTDESGTQIRIGTAPVETIKYKQOT 404

RESULT 9

08VW87 PRELIMINARY: PRT: 578 AA.
AC 08VW87.
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Beta-fructofuranosidase.
GN BFF.
OS Arthrobacter sp. K-1.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococciaceae; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=160970;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K-1;
RA Ito T., Fujita K., Hara K., Tomozuka T., Sakano Y.;
RT "Molecular Cloning of Beta-fructofuranosidase from Arthrobacter sp. K-1";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB062134; BAB72022.1; -
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
DR SEQUENCE 578 AA; 62839 MW; BCEACE34E3F02F0B CRC64;

Query Match 11.3%; Score 278.5; DB 2; Length 578;
Best Local Similarity 23.7%; Pred. No. 1.4e-09;
Matches 125; Conservative 67; Mismatches 193; Indels 143; Gaps 25;

0Y 8 EDYGFATHTRADMLKI-----PGQONSPOFK-----VPOFNASAIKNDISAKGYDKS 54

DB 57 EDDFTAKTTRADARQLOKMSDPTAPSRENSMVASVTMPVTPQ----- 98

55 GNLIDLV-----VWDSPLQNDGTAANYHGHIVSALAGDK-NSDD-----TPLLHLYQ 104
99 -DEPMSNEQVWVWMTPLTDEDANQSVNGEIIIFSLVADBNLGFDRHVPRAKIGYEYR 157
105 KVGDSID-----SWKNAGRVF-EDMDKFPVNDPLKTYOTOMSGSATLTGQVRLFTY 158
158 PAVGPAEREPENGMTYGGVLFKEGYTGOIFEDQSFHOTO-WSGSARVSKNGELTKLFT 216
159 DYSGNPEDEGTGA-----GNOIISTAQVNLSDPDAATLVKDVSDKSVFDGSDGYVONI 214
217 DVAFYRNSDGTNIKRYDPRIALSVGVKANKGVTLLTGKNTKTLTLQ-----ADGYTYQTG 272
215 QQFIDEGKWSGDNHTLRDPHYVEKGH---KYLVEAN-----TGTTD---GY-QGD 260
273 AQ-----NEFNFRRDPFTEEDPAHGFETPMVEGNSAMQRETAATCNEADLGYROD 323
261 QSFNNKAYVGGSDVDFQNEKNKLLQSPKQOIASLANGALGIYELADDTYKVSVMPLVAS 320
324 PYAETVDDVNASGATYQIGNGLAKAKKOL-----TEWFLPPLISA 366
321 NTVADEVERANIEKNNKWTLETDGSGKMTSDGINDKDYMLGPGSLNGPNNPI-E 379
367 NCVTDQTERPQLYEFKDKSYLFTTISHRGTFAGLDGPEGV--GVYGDGIRSDYQPLNG 424
380 TGLVNLMLDPAIDL-----HTYSHGCIHPHEGNNVLTYSMTNRGF 421
423 SGLALG---NPTNLFGLGQFPAPDFNDHGFHQAYSHYVMP-----GGLVQSFIDTIG- 475
422 YPEHSHLKD-----KLGVNIKSDTS-----GGNSSGQGOFP 455
476 --THDFVAGGTLPATPVKMDIGVGDPTKTAVDYSGSGLGGMADIP 521

RESULT 10

09LXL1 PRELIMINARY: PRT: 430 AA.
AC 09LXL1.
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Levansucrase.
GN LSXA.
OS Acetobacter xylinum.
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=28448;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NC11005;
RX MEDLINE=20450679; PubMed=10997873;
RA Tajima K., Tani T., Kobayashi Y., Kohno H., Fujiwara M., Shiba T.,
RA Erata T., Muneoka M., Takai M.;
RT "Cloning and sequencing of levansucrase gene from Acetobacter xylinum
NCI 1005";
RL DNA Res. 7:237-242(2000).
DR EMBL: AB034152; BAA93720.1; -
DR InterPro: IPR003469; Glyco_hydro_68.
DR Pfam: PF02435; Glyco_hydro_68; 1.
DR SEQUENCE 430 AA; 47476 MW; 7EEB80C2F521100C CRC64;

Query Match 9.8%; Score 240.5; DB 2; Length 430;
Best Local Similarity 23.7%; Pred. No. 2.3e-07;
Matches 106; Conservative 60; Mismatches 135; Indels 147; Gaps 23;

0Y 13 AHITRADMLKIPGOONSPOFVPOPNASAIKNDIAKGYDKSGNLIDLV--WDSMPLO 70
DB 16 SHMTTADAMKAVIADPTTTPYIDYNF-----YIDQVWQMDGSLRA 59
0Y 71 ADGTANYHGHIVSALAGDPKNSDTP-----LHLFYOKVGDTSIDSKMNGR 119

DB 60 ITGETVKENDMYVMALVANRADTGDTEGWHNRNFAVIGFYSRNG---ID-WTFGGR 115
 QY 120 VEDEDKDFVNDPRLKYQTQEWSSATL---TKDQVRLFTYDYGNDPEDGTGAGNOII 176
 DB 116 LIQKADLRFPD-----EMSGSLVIRACTKN-TVDPEFTSVN----- 150
 QY 177 STAQVNLQOPDPAATLKVDGVSDDKSVFDG-----GDGTVYONIQ--QFIDEGKWTIS 225
 DB 151 --TDINOSVPSGCTKIFS-NDEKVMFEGFSKTVEMFASDGVNANAEEDQYFD----- 201
 QY 226 GDNHRLRDPHYED--KGHKYLVEFANGTGTYGDDGSFNKK-----AVY 269
 DB 202 ----FRDPHPLNADPKIKYCLFE---GNVPVGKGFLLHDAIGAVPPGYTPAGAGY 253
 QY 270 GGSDFVQNEKNKLQSPKQIASLANGALGIVELADDTYVK-----SYMKPLVASN 321
 DB 254 GAA-----AIGIARLADGAYEKGDSSRHMTLLPALVAL 288
 QY 322 TVADEVERANIFKMNKKYLFDSRSGSKWTSDDGINDKDYV-MLGPGGDSLNGPHNPINET 380
 DB 289 GVNDDQTERPHYFKNDNTYIFTISHSTYTGDSGTGPDGYGFSVSENG--IFGYPBPLNAS 346
 QY 381 GLVLMNNDPADLTH-TYSHGCIHPREG 407
 DB 347 GLVLDG--NPSAPYETYSH--FVDPDG 369

RESULT 11

Q932F7 PRELIMINARY: PRT: 1141 AA.

AC 0932F7
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding
 DE protein.
 GN SDRF OR SAV0563.
 OS Staphylococcus aureus (strain Mu50./ ATCC 700699).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Staphylococcus.
 OX NCBI_TaxId=158878;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21311952; PubMed-11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iken J.-O., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Iabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 aureus";
 RT Lancet 357:1225-1240(2001).
 RL EMBL, AP003355; BAB56725.1.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor.1.
 DR TIGRFAMs: TIGR01167; LPTYG_anchor.1.
 DR TIGRFAMs: TIGR01168; YSIRK_signal.1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 1141 AA; 124038 MW; E679F7C2991846D9 CRC64;

Query Match

Best Local Similarity 5.7%; Score 141; DB 16; Length 1141;
 Matches 119; Conservative 61; Mismatches 198; Indels 170; Gaps 26;

QY 3 SGDKYEDYGFAMHTRADM-----KIPQONSPOKFPVQFN-----ASAINKIDSAR 49
 DB 464 AGSQVDYDGNIKLNGSTIIDQNTFIRKYKVNPNQQLPQSNIRYDFSQYEDVTSGDFMK 523
 QY 50 GYDSGNLIDDVMD-----SWPLQVNDGTA-----NHGCHI 83

DB 524 SF--SNNVATLDFGDIINSAYIIKVVSKYTPISDGLDIAQGTSMRTTDKGYVNYAGYSN 581
 QY 84 VSALAGDKRNSDDT--PLHLEFYQKVGDTSIDSKNAGVFPDMOKFVNDPRLKYQTQEW 141
 DB 582 FLTVISNDGGGGGTAKPEPKLY-KIGD-----YWEDEVDR---DVGQIDTSKEK 626
 QY 142 SGS---ATLT--KDQVRLFTYDYGNDPEDGTGAGNOIISTAQVNLQOPDA-ATLKVDGV 196
 DB 627 PMANLVLVLTYPDGTGKSVRTDANANGHEFGGLKDS---EYTYVFEFPAVLPKVNKT 682
 QY 197 SDHKSVEFGGDTYQNIQO-----FIDEGKWTISGNNHRLRDPHYEDKGHKYLVFA 249
 DB 683 TDGEKDSNGSSITVYKINKDMSLDPTGFYKPEKYNLGD-----YWEEDTNKDGIDQ 734
 QY 250 NNGTTDVGQDGSFNKAYVGGSDVFPQNEKNKLQSPKQIASLANGALGIVELAD--- 306
 DB 735 NEP-----GIKDV-----KYLKDSIGKAYIGTTTIDASKRYKFTDLDN 772
 QY 307 -DYTVK-----SYMKPLVASNTVAD-----VERANIFKMN-----K 338
 DB 773 GNYTYFEFEPFAGYPTVYKNTTAEDKDSNGLTGTVIKDADNMTLDSGFYKTPKYSLDGYV 832
 QY 339 WYLFDSGSKWTSDDGINDKDYVMLGPGGDSLNGPHNPINETGLVLMNLD----- 389.
 DB 833 WYDSNKKDKQDSTEGIDVAVKVTLLNEKEVIG--TTKIDNGKRYFPDNLDSGKRYVFE 890
 QY 390 -PADLFTTYSHC--GIPPEGNNVLT-----SYMTNGRFPYEHSHLRDLKGVNIKSD 441
 DB 891 KPAGLTQTVTNTTEDKRDADGGEVDVTTTDDHDFLIDNGYFEE-----D 934
 QY 442 TSGGENSS 449
 DB 935 TSDSDSDS 942

RESULT 12

Q9F3Z5 PRELIMINARY: PRT: 1449 AA.

AC Q9F3Z5
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Adhesion penetration protein.
 GN App.
 OS Neisseria meningitidis.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxId=487;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-23515;
 RA Ala-Aldeen D.A.;
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-23515;
 RA Abdel Hadi H., Woolridge K.G., Robinson K., Ala Aldeen D.A.;
 RT "Identification and Characterisation of App: an Immunogenic
 Autotransporter Protein of Neisseria meningitidis";
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AJ296276; CAC14670.1.
 DR MEROPS: S06.006;
 DR InterPro: IPR000710; IGA_S6.
 DR InterPro: IPR004899; Pertact_sup.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR Pfam: PF03212; Pertactin.1.
 DR Pfam: PF00921; IGASERPTASE.
 DR PRINTS: PM00020; Tryp_Spc.1.
 DR SMART: SM00020; Tryp_Spc.1.
 DR PROSITE: PS00135; TRYP_SIN_SER; UNKNOWN_1.
 KW Hydrolyase: Serine protease.
 SQ SEQUENCE 1449 AA; 159071 MW; C07850BD4B7C25D CRC64;

Query Match

5.7%; Score 141; DB 2; Length 1449;

DR Pfam: PF00746; Gram_pos_anchor: 1.
 DR TIGRfams: TIGR01167; LpxTG_anchor: 1.
 DR TIGRfams: TIGR01168; YsirK_signal: 1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 1141 AA; 124026 MW; 44541900B8C5A4F8 CRC64;

Query Match 5.7%; Score 139; DB 16; Length 1141;
 Best Local Similarity 21.7%; Pred. No. 2.1;
 Matches 119; Conservative 60; Mismatches 199; Indels 170; Gaps 26;

```

OY 3 SGDKEDYGFHITRADML-----KIPQONSPOKVPQFN-----ASAINKIDSAR 49
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 464 AGSQVDYDGNIKLNGSTIIIDONTETIKYKVPNQQLPQSNRIYDSQYEDVTSQDNKK 523
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 50 GYDKSGNLIDLDVWD-----SMPLOMADGTAA-----NHGYHI 83
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 524 SF--SNVATLDFGDINSAYIIKVVSKYTPSDGLDIAOGTSMRTDKYGYNYAGYSN 581
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 84 VSALAGDPKNSDDT--PLHLFYQKVGDTSIDSMKNAGRVFEDMDKFPVNDPYLKYOTQEW 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 582 FIVTSNDTGGGDTYKPEEKLY-KIGD-----YVWEDVDK----DGVQGTDSKEK 626
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 142 SGS---ATLT-KDQVRLFTDYSGNPEDEGTGAGNOIISTAQVNLSPDA-ATLKVDGV 196
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 627 PMANVLVLTLYPDGTTKSVRTDANGHYEFGGLKDG---EYTVKFEETPGYLPRTKVGNT 682
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 197 SDHKSVFPGDGTGYQNIQO-----FIDEGKWIISDNHTLRDPHYEDKGHKYLVEA 249
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 683 TDGEKDSNGSSITVYINKDKDMSLDTGFKYKRYMLGD-----YVWEDTNKIGTIDA 734
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 250 NTGTDTGYQGDQSFNNKAYYGSDVFQNEKNKLLQSPKKOIASLANGALGIVELAD-- 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 735 NEP-----GIKDY-----KVTLKDSGTGKVIIGTTTIDASGKRYKFTLDN 772
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 307 -DYTVK---SYMKPVLASNTVADE-----VEKANIIFKMN-----K 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 773 GNYTVETFPAGYTPVYKNTTAEDKDSNGLTGTTGVIKADNMNTLDSGYYKTPKYSLGDV 832
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 339 WYLFDSRSGSKMTSDGINDKDVYMLPGGDSLNGPHNPINETGLVNNMLD----- 389
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 833 WYDSKDKKQDSTENGKIDVYTLNKEGEVIG--TTTIDENGKRFNDLSGKRYLVE 890
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 390 -PADLFTHTYSHC--GIRPPEGNNVLT-----SYMTRNGFYPEHSHLRDLKGVNIKSD 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 891 KPAGLTQTVYNTTEDDKDADGGEVDVTTIDHDFTLIDNGYFEE-----D 934
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 442 TSGGENSS 449
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 935 TSDDSDSDS 942
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

ID 086489 PRELIMINARY: PRT: 1166 AA.
 AC 086489;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Sdr E protein.
 GN SDR E.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NEWMAN;
 RX MEDLINE=99098700; PubMed=9884231;
 RA Josefsson E., McCrea K., Ni Eidhin D., O'Connell D., Cox J., Hook M.,
 RA Foster T.J.;
 RT "Three new members of the serine-aspartate repeat protein multigene
 family of Staphylococcus aureus.";

RL Microbiology 144:3387-3395(1998).
 DR EMBL: AJ005647; CA00652.1; -.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor: 1.
 DR TIGRfams: TIGR01167; LpxTG_anchor: 1.
 DR TIGRfams: TIGR01168; YsirK_signal: 1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 SQ SEQUENCE 1166 AA; 126548 MW; 750A7B0135287DA4 CRC64;

Query Match 5.6%; Score 138; DB 2; Length 1166;
 Best Local Similarity 22.2%; Pred. No. 2.5;
 Matches 122; Conservative 53; Mismatches 201; Indels 174; Gaps 28;

```

OY 3 SGDKEDYGFHITRADMLKIPQONSPO-----KVPQFNASAINKIDSAR 48
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 469 AGSQVDYDGNIKLNGSTIIIDONTETIKYKVPNQQLPQSNRIYDSQYEDVTSQDNKK 527
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 49 GYDKSGNLIDLDVWD-----SMPLOMADGTAA-----NHGYHI 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 528 KSF--SNVATLDFGDINSAYIIKVVSKYTPSDGLDIAOGTSMRTDKYGYNYAGYS 585
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 83 IVSALAGDPKNSDDT--PLHLFYQKVGDTSIDSMKNAGRVFEDMDKFPVNDPYLKYOTQEW 140
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 586 NFVTSNDTGGGDTYKPEEKLY-KIGD-----YVWEDVDK----DGVQGTDSKEK 630
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 141 WSGS---ATLT-KDQVRLFTDYSGNPEDEGTGAGNOIISTAQVNLSPDA-ATLKVDG 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 631 KPMANVLVLTLYPDGTTKSVRTDANGHYEFGGLKDG---EYTVKFEETPGYLPRTKVGNT 686
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 196 VSDHKSVFPGDGTGYQNIQO-----FIDEGKWIISDNHTLRDPHYEDKGHKYLVE 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 687 TDGEKDSNGSSITVYINKDKDMSLDTGFKYKRYMLGD-----YVWEDTNKIGTIDA 738
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 249 ANTGTDTGYQGDQSFNNKAYYGSDVFQNEKNKLLQSPKKOIASLANGALGIVELAD-- 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 739 ANEP-----GIKDY-----KVTLKDSGTGKVIIGTTTIDASGKRYKFTLD 776
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 307 -DYTVK---SYMKPVLASNTVADEVERANI-----FKMNK----- 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 777 NGNYTVETFPAGYTPVYKNTTAEDKDSNGLTGTTGVIKADNMNTLDSGYYKTPKYSLG 835
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 339 --WYLFDSRSGSKMTSDGINDKDVYMLPGGDSLNGPHNPINETGLVNNMLD----- 389
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 836 YWYDSNKNQDQSTENGKIDVYTLNKEGEVIG--TTTIDENGKRFNDLSGKRYLVE 893
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 390 -PADLFTHTYSHC--GIRPPEGNNVLT-----SYMTRNGFYPEHSHLRDLKGVNIKG 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 894 FEKPAGLTQTVYNTTEDDKDADGGEVDVTTIDHDFTLIDNGYFEE----- 938
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 440 SDTSGGENSS 449
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 939 -DTSDSDSDS 947
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: June 16, 2003, 12:09:40
 Job time : 117.682 secs